2013

The Impact of Data Complexity on Privacy Management Resulting from Vehicular (V2X) Applications

Andre Zierfuss
University of Rhode Island, a.zierfuss.uri@gmail.com

THE IMPACT OF DATA COMPLEXITY ON PRIVACY MANAGEMENT RESULTING FROM VEHICULAR (V2X) APPLICATIONS

BY

ANDRE ZIERFUSS

A THESIS IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

UNIVERSITY OF RHODE ISLAND

2013

MASTER OF SCIENCE

OF

ANDRE ZIERFUSS

APPROVED:

Thesis Committee:

Major Professor Resit Sendag

Joan M. Peckham

Haibo He

Qing Yang

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2013

ABSTRACT

In recent years, privacy management has become one of the most complex processes in the connected world. Fundamental technologies like GPS, cellular communications, and the Internet have become mandatory equipment in the modern vehicle. Subsequently, the vehicle became part of this connected world, wherein data are constantly sent and received. Accordingly, it became inevitable to introduce data security to vehicular communication. Hence, the development of location based and other connected services, introduced a new level of data complexity. In scenarios where GPS data are tied to certain entities or databases consisting of entire personal profiles, data cannot be treated separately anymore. Prior improvements regarding privacy protection achieved through anonymous pseudonyms have become negligible, due to GPS enabled traceability. This paper presents a new approach that turns privacy protection from a one-way street into a negotiation process. It allows the user to individually decide what data is provided and what is kept private.

ACKNOWLEDGMENTS

I owe much gratitude to my URI advisor, Resit Sendag, for his guidance and support especially in the final steps of my thesis. I also want thank my ERL advisor Stefan Sellschopp for always trying to give guiding input, and pushing my work in the direction of success, wherever he could. Both advisors have taught me different but complementary knowledge that I will always cherish throughout my engineering career. I am grateful for having the chance to work at Volkswagen Electronics Research Laboratory in the heart of the Silicon Valley. It has been a great experience that allowed me to meet lots of different engineers with inspiring ideas. I thank my other thesis committee members: Joan M. Peckham, Haibo He and Quing Yang, for making their time and knowledge accessible to me. I owe a lot to Simon Kwoczek and Ananth Raghunathan for their advice and time.

I wish to also thank my TU Braunschweig advisor Wael Adi, who chose to work with me once again and took his time to advise me, where he could. I am very grateful to Thomas Form for bringing me in touch with the ERL in the first place and to Stefan and Simon for taking me into their team. It has been an exceptional experience that helped me not only to evolve professionally but also in terms of self-organization and self-assessment that will help me attack my new challenges. Last but far from least, I am forever indebted to my family and friends for their love and support. I have been blessed with a very special family and I can tell there is no greater support than knowing that someone is always close to you, no matter you live. My graduate studies have been quite a journey and I cannot wait for the next one.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGMENTS iii

TABLE OF CONTENTS iv

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER 1 – INTRODUCTION 1

MOTIVATION 3

CONTRIBUTIONS AND FINDINGS 5

OUTLINE DESCRIPTION 8

CHAPTER 2 – PRIVACY AND ITS MANAGEMENT FOR CONNECTED VEHICLES 9

AUTOMOTIVE COMMUNICATIONS 9

Automotive Telematics 9

Vehicle-to-Infrastructure (V2I) Communications 11

Vehicle-to-Vehicle (V2V) Communications 12

Social Media Communication 13

Automotive application data categories 13

PRIVACY PROTECTION POLOCIES 15

Privacy definitions 15

U.S. vs. European governmental privacy policies 15

Golden Rules 18

GENERAL DATA AND IT-SECURITY MECHANISMS 19

Certificates 19

Cipher methodologies 24

CHAPTER 3 – VEHICLE DATA EXTRACTION AND DISTRIBUTION FRAMEWORK 34

INTRODUCING THE FRAMEWORK 34

DATA PROCESSING 35

POSSIBLE CONCEPT ALTERATIONS 36

CHAPTER 4 – PRIVACY PROTECTION INTEGRATION 39

HOW PRIVACY PROTECTION MECHANISMS VARY WITH THE APPLICATION 39

PRIVACY PROTECTION CONCEPT FOR THE PRESETNED FRAMEWORK 40

PRIVACY PROTECTION CONCEPT FOR THE HOTSPOT IDENTIFACION PROTOCOL 51

Protocol overview 51

Identified privacy protecting mechanisms in the protocol 53

CHAPTER 5 – IMPLEMENTATION 61

IMPLEMENTING THE PRESENTED FRAMEWORK 61

SENDING SAMPLE IS SENT FROM THE VEHICLE TO THE BACKEND 66

CHAPTER 6 – EXPERIMENTAL SETUP 72

IMPLEMENTATION TEST CASES 72

GOLDEN RULES 75

CHAPTER 7 – EVALUATION 77

TESTING THE COMMUNICATION 77
DISCUSSING THE RESPONSIBILTY OF PRIVACY PROTECTION 82
PRIVACY PROTECTION MECHANISMS APPLIED TO THE MAIN AND THE RELATED CONCEPTS 84
Fulfill the privacy protection mechanisms the required policies? 84
PRIVATE HOTSPOT IDENTIFICATION AS A DIFFERENT DATA DISTRIBUTION CONCEPT 87
Fulfill the privacy protection mechanisms the defined policies? 87
DERIVED DATA PRIVACY CLASSES 92
CHAPTER 8 – RELATED WORK 94
PRIVACY PROTECTION THROUGH ANONYMITY 94
PRIVACY PROFILE MANAGEMENT 97
LATEST PRIVACY DISCUSSIONS 99
The right to be forgotten 99
Current privacy cases regarding traveling safety 100
Current privacy cases regarding social media and connected applications 101
CHAPTER 9 – CONCLUSION 102
BIBLIOGRAPHY 105

LIST OF TABLES

TABLE PAGE

Table 1: Telematics data application overview 10

Table 2: Vehicle-to-infrastructure data applications overview 12

Table 3: Vehicle-to-vehicle data applications overview 13

Table 4: Social data applications overview 13

Table 5: Application data categories 14

Table 6: Certificate parameters 21

Table 7: Profile variations for a travel guide service 49

Table 8: Multiplicative inverse of (d) in $\mathbb{Z}_{132}$ 58

Table 9: Multiplicative inverse of (r) in $\mathbb{Z}_{161}$ 59

Table 10: API documentation overview 67

Table 11: Defined Test Cases 73

Table 12: Defined Inputs and Outputs 74

Table 13: Test cases including success criteria 77

Table 14: Policies and Responsibilities 82

Table 15: Policies and mechanisms regarding the new framework 85

Table 16: Policies and mechanisms regarding the hotspot protocol 89

Table 17: Privacy level based on defined data categories 92

LIST OF FIGURES

FIGURE PAGE

Figure 1: Digital signing 19
Figure 2: Digital verification..... 20
Figure 3: Certification management..... 22
Figure 4: Symmetric cipher algorithm 24
Figure 5: Data Encryption Standard (DES) with one key (K) encryption 25
Figure 6: Data Encryption Standard (DES) with (N) key (K) encryption 25
Figure 7: Asymmetric cipher algorithm..... 26
Figure 8: Data processing..... 35
Figure 9: Protocol data processing concept 37
Figure 10: Protocol Architecture..... 37
Figure 11: Privacy enhanced framework 46
Figure 12: Hotspot identification registration protocol..... 52
Figure 13: Hotspot identification voting protocol..... 52
Figure 14: Software overview of the implemented framework 61
Figure 15: RESTlet client-server architecture..... 63
Figure 16: Hierarchal representation of the JSON API 64
Figure 17: Physical communication in Kafka..... 65
Figure 18: Data translation overview..... 66
Figure 19: Conceptual system overview..... 72
Figure 20: JSON representation of the vehicle data..... 78
Figure 21: XML source of the vehicle data 78
Figure 22: First part of the update..... 79

Figure 23: Second part of the update 79
Figure 24: Timestamp and extracted vehicle data..... 80
Figure 25: Extracted vehicle data..... 80
Figure 26: First subscription 81
Figure 27: Second subscription..... 81
Figure 28: Data protection responsibility breakdown..... 83
Figure 29: Kalman iterations..... 95
Figure 30: Kalman based muli hypothesis tracking..... 95
Figure 31: Variation of beaconing intervals..... 96
Figure 32: IBM MyPrivacy Component Architecture 97

CHAPTER 1 – INTRODUCTION

This thesis presents a privacy management concept that allows for privacy protection, while incorporating the latest developments in data complexity regarding vehicular applications. In general, location based services and services that involve location information, have significantly gained popularity among users and car manufacturers. The challenge at this point is, whenever general data, such as status information or location information, is combined with sensitive information like personal data, this new individual driving data cluster becomes sensitive as well. In other words, an increase in data complexity also results in an increase in security complexity (Fiaschetti et al., 2012). In order to protect these clusters, new data security classes need to be determined based upon the need for physical storage protection, access control, and required protection level.

In previous work it has been shown that state of the art pseudonymous communication does not guarantee the required privacy, due to traceability (Wiedersheim et al., 2010). The purpose of this thesis is to provide an overall data analysis that allows for data classifications including correlating privacy classes. These classes shall be defining the basis for an automotive privacy model improving the flexibility and transparency of data protection based on the *IBM-My Privacy Component Architecture*, originally introduced for the Internet by Bohrer et al. in 2001. A live vehicle-data extraction and distribution framework shall be utilized to identify the relevant data clusters and to evaluate and demonstrate the need for state of the art security mechanisms to be applied to different automotive use-cases.

Accordingly, it shall be shown how privacy protection mechanisms vary with the targeted use-case. Further, the general breakdown of responsibility regarding privacy protection among the involved developing parties shall be discussed. Additionally an extract of ongoing privacy debates shall illustrate how current laws, especially in the U.S., do not provide the necessary legal framework to protect personal data.

In the past the focus of security solutions was related to in-vehicle security (Schweppe et al., 2012). Most of the data stayed either within the vehicle or were pulled from an external source (points of interests - POIs) that had no immediate relation with customers' driving data. Accordingly, data monitoring was only necessary one way (intrusion detection). Now that data become steadily more individual and the back-end communication increases as vehicles become also more traceable, the protection of the driver's privacy becomes more complex as well. Besides traceability, storage protection and access control have become even more crucial in order to protect each entity within the communication process.

The result of this thesis shall deliver a classification model of the major data clusters generated from automotive applications and derived data security classes as described above. These classes shall serve as basis for the introduced privacy concept.

In this chapter, the motivation is to briefly present the work in this thesis, regarding why pseudonyms do not offer full anonymity, the current privacy laws in the U.S. are not yet prepared for private data collections, and what general changes in privacy management are necessary in order to offer more transparency and flexibility to the customer.

MOTIVATION

For a lot of inter-vehicle communication (IVC) system services, GPS data represent a major asset to various functions. Global position data is fundamental for location based services like recommendations for specific routes (e.g. scenic versus most direct routes), social events (hotspot identification), requesting the nearest business or service (e.g. ATM or restaurant), or turn-by-turn navigation to any address.

It is obvious that IVC systems work with and thereby reveal very detailed location information patterns about the vehicle. A common and widely accepted security mechanism is the use of pseudonyms allowing to anonymously authenticating identities. However, the mapping of any kind of data with very precise GPS data allows generating a very detailed personal picture of the driver. In General, Wiedersheim et al. have demonstrated the possibility of reconstructing long traces of a majority of vehicles within the same area. According to their work it is more than questionable if location privacy is achievable in IVC systems against a powerful adversary. Even though actual identities are replaced by pseudonyms and those also change over time, once a target is identified based on its location every vehicle can be tracked. The attempt to change the location data density has not yielded the desired results due to standardization constraints (for more details see *Chapter 8*).

Another recent approach regarding privacy protection was cutting out any kind personal data and thereby reducing the information exchange to simple quantity statements. One example is the private identification of location hotspots introduced

by Raghunathan et al. The protocol provides an anonymous voting mechanism that detects location hotspots, while not revealing anything about the voting participants. It defines optimal privacy protection.

In contrast, services like online diagnosis, route recommendations or charging recommendations for electric vehicles require much more detailed data to be exchanged. In other words, in order to make use of these services at some point, data needs to be exposed. The popularity of mobile application has revealed that users are willing to provide certain data in order to make use of the various data services. As already mentioned in the introduction, one goal of this thesis is to provide a more flexible and also more transparent concept for privacy protection. The key to widely accepted services is transparency. As an example, the majority of smart phone users nowadays receive notifications when applications request access to certain kinds of data. The same transparent application profile management that integrates the user's decision can easily be applied to the automotive environment. This thesis will discuss both mentioned data concepts regarding privacy protection and usability.

The ubiquitous topic of privacy protection has led to major discussions among governmental parties, influential companies, and independent privacy authorities. Answering the question of how to protect private data appropriately has now been discussed for decades. It appears to be that the European Union has taken the leading role in these controversial debates, whereas the United States only started very recently to pay attention to privacy concerns. The latest privacy discussion among the authorities involves the integration of a communication device, also known as "black box" (as it exists in planes; Lowy, J., 2012). According to a governmental decision,

the device will be integrated into every new manufactured vehicle, for safety reasons. Privacy authorities claim that there exist no rules or policies that define limitations to the data collection enabled through this device. These conditions demand immediate action in order to protect privacy, this thesis will incorporate the following two privacy protection aspects. The first is to define specific privacy requirements based on the already mentioned hotspot detection protocol that provides optimal privacy protection and shall therefore be used a guideline. Second, the privacy protection directive given by the European Union is a worldwide-accepted directive and shall serve in this thesis as privacy protection policy standard.

CONTRIBUTIONS AND FINDINGS

This thesis has implemented a live vehicle-data extraction and distribution framework. So far, data extraction has been available in many ways for in-vehicle data that is intended to be further processed on a back-end server. The current sensor data of a vehicle is mostly exchanged among the various electronic control units (ECUs) that communicate over the main in-vehicle network, the controller area network (CAN) bus (ISO 15765-2, 2004). Alternatively, greater amounts of data can be sent to a server that runs more complex data processing algorithms. These algorithms can either be used for internal purposes or to offer specific services like, for example route recommendations based on the driving style and vehicle model. The framework presented in this thesis provides a flexible subscription to all the available data (i.e. mostly CAN data) that can be used in various data services. Every service can decide

individually, through specific subscriptions, instead of simply tapping into a bus network that is flooded with all the available data. Additionally, the communication between the vehicle and the back-end is established over a wireless connection, which allows for even more flexibility. The communication build into the vehicle also incorporates GPS data, which is needed for the recommendation of charging-friendly, fun or scenic routes close to the customer.

As already mentioned, whenever GPS data is involved, privacy becomes a major concern. One of the latest privacy protecting approaches focuses on the exchange of less data by cutting out any kind personal data. This concept provides clearly optimal privacy protection but significantly decreases the service opportunities. This thesis presents a privacy concept that adapts the needs the of detailed data services, while applying similar privacy requirements presented in the previous approach, but also allows for more flexibility. Besides state of the art security mechanisms are protecting the general communication, a transparent application profile management shall be integrated. This profile is inspired by the IBM My Privacy Component Architecture. It provides several primary components based on the IBM's Enterprise Privacy Architecture (EPA) handling privacy concerns originally intended for the Internet. This thesis shows that this concept can easily be integrated into the automotive environment, with requirements derived from the "Golden Rules". The Golden Rules are part of the German Federal Data Protection Act, the so called "Golden Rules." These Golden Rules are also mounted into the British and most other European country's federal laws. The act focuses on the protection against misuse of personal data in terms of data processing (BDGS, 1994, also see *Chapter 2 and 6*). The profile

management shall improve privacy protection, by allowing the vehicle client to decide, what data shall be available for subscription.

Both the framework and the privacy concept have been evaluated. Test cases for the framework have been defined based on standard black-box and white-box tests, looking at inputs and outputs, and whether the internal data processing operates as required. The state of the art privacy mechanisms applied in the guiding hotspot identification concept and the newly defined live vehicle-data extraction and distribution concept have been validated through mathematical proof of concept. Additionally the privacy concepts discussed, have been validated against the privacy requirements stated by the Golden Rules.

As a basis for the privacy profile management, this thesis gives an overview of the various automotive applications and assigns data security classes derived from the IBM concept and translated into the automotive context.

OUTLINE DESCRIPTION

The structure of this thesis is divided into nine fundamental parts:

Chapter two gives an overview of current data applications and available privacy policies and security mechanisms

Chapter three introduces the use-case definition, and the live vehicle-data extraction and distribution framework

Chapter four describes a conceptual integration of the created privacy concept

Chapter five gives an overview of the framework implementation

Chapter six provides the experimental setup to evaluate the introduced framework and the related privacy concept

Chapter seven presents the evaluation results of experimental setups

Chapter eight gives an overview on related work including common data application implementations, privacy concepts and current legal cases concerning privacy

Chapter nine gives an overall conclusion summarizing the work and findings of this thesis

CHAPTER 2 – PRIVACY AND ITS MANAGEMENT FOR CONNECTED VEHICLES

This section will introduce the most common vehicle-to-x and other connected user applications. In a second step, the available applications shall be summarized data categories. Eventually, state of the art privacy protection policies and mechanisms shall be presented.

AUTOMOTIVE COMMUNICATIONS

Automotive Telematics

The fundamental technologies for automotive implementations were first aligned in the 1990s with GPS, cell phone technology and the Internet (Cordis, 2013).

The most precise definition of telematics is, "a wireless communication system designed for the collection and dissemination of information that particularly refers to vehicle-based electronic systems; vehicle tracking and positioning; online navigation; and information systems and emergency assistance" (Tutorials point, 2013).

The telematics system is implemented into the vehicle as the telematics communications unit (TCU) that communicates wireless with a central service center. The TCU functions as a central platform of the vehicle telematics system that incorporates all telematics-associated technologies. It provides location-specific information to a central service center, whereas the center helps to deliver telematics services to a vehicle via cellular phone. Further, the TCU is linked to the engine

control unit (ECU) allowing for enhanced services i.e. remote engine diagnostics and automatic airbag notification. A very common application that has been integrated into the vehicle for years is related to a system called "black-box" (as it exists in planes; Lowy, J., 2012). Before signals are transmitted as wireless signals they will be collected through this module, which is placed behind the dashboard. It integrates a phone, GPS receiver, digital signal processor and microphone for voice recognition. Additionally it incorporates the vehicle's data bus to collect diagnostic information from the available sensors. On the other side, the so called back-end server functions like an Internet server, i.e. handling applications (analyzing diagnosis data) (Electronics-TCU, 2013). Other current and future telematics applications are listed below (Tab.1).

Table 1: Telematics data application overview

Application	***Description***
Navigation services (POIs)	Provides extended navigation services
Web radio	Provides access to online radio
Wi-Fi hotspots	Provides Internet hotspot functionalities
Traveling information,	Provides specific information about the area
Weather	Provides weather forecasts and related information
Nearest gas/charging station	Provides closest gas/charging station near your current location
Emergency assistance	Allows for emergency calls from the vehicle, providing position and other information

Recommendation engines	Provides different recommendations according to different behaviors
Online diagnosis	Provides immediate simple maintenance, and sets up needed appointments
Pay-for-use Insurance	Provides an insurance rate that is based on the driving behavior
Black box systems	Provides crash monitoring (officially from Sep. 1st 2014)
Back2car	Provides vehicle fleet management and extended connected services

Vehicle-to-Infrastructure (V2I) Communications

Vehicle-to-infrastructure communications enable data transmissions from the roadside to the vehicle in order to alert drivers that e.g. it is not safe to enter a certain intersection. Data can be exchanged by using vehicles as data collectors, to anonymously transmit traffic and road condition data from all main roads of the transportation network. These data help to provide transportation agencies with the knowledge that is needed to implement dynamic plans allowing for reduced traffic congestions. One well-known application in vehicle-to-infrastructure communication is the electronic toll collection system. It is present in many countries and therefore somewhat advanced in terms of deployment. As already mentioned, other applications that shall enable the reduction of vehicular accidents, traffic congestions, transportation time, fuel consumption and environmental impact of road transport,

define the fundamental research areas (Holfelder et al., 2004; Jansons et al., 2012, National VII Coalition, 2013).An overview of current and future V2I applications is given below (Tab.2).

Table 2: Vehicle-to-infrastructure data applications overview

Application	***Description***
Wireless payment or toll systems	Provides wireless payment systems at toll stations, (gas stations and parking garages)
Real time traffic information	Provides real time traffic within the network
Traffic light assistant	Provides congestions control through the distribution of traffic light timings

Vehicle-to-Vehicle (V2V) Communications

For many years now, vehicle-to-vehicle communication has been addressed by automotive and non-automotive research organizations. The most common use cases include numerous infotainment applications i.e. ad-hoc networking for information exchange, chat applications or gaming, but also advanced active safety applications i.e. inter-vehicle hazard warning or spectrum intersection collision avoidance systems. Until today, not any of those applications have moved to production due to technical and non-technical i.e. business related issues (Holfelder et al., 2004). An overview of current and future V2V applications is given below (Tab.3).

Table 3: Vehicle-to-vehicle data applications overview

Application	***Description***
Accident reporting	Provides accident report distribution
Warnings (entering intersections, departing highways, sudden halts, lane change)	Provides context-based warnings
Adaptive cruise control (ACC)	Provides automatic vehicle speed adjustments

Social Media Communication

According to current development trends, social platforms like Facebook and other social applications like Twitter are planned to be integrated into the modern vehicle's infotainment system (Tab.4).

Table 4: Social data applications overview

Application	***Description***
Facebook	Provides access to friends activities
Twitter	Provides access to friends activities and news

Automotive application data categories

The table below (Tab.5) shows the prior introduced data applications grouped in three data categories as a basis for the later evaluation that shall result in the required data security classes.

Table 5: Application data categories

Application type	*Identifiable data*	*Time/location based data*	*Broadcasted data, or independent data*
V2I	Wireless payment or toll systems	real time traffic information	Traffic light assistant
V2V		Accident reporting	Warnings (entering intersections, departing highways, sudden halts, lane change) Adaptive cruise control (ACC)
Telematics	Online diagnosis	Navigation services (POIs)	Web radio
	Wi-Fi hotspots Pay-for-use Insurance Emergency assistance	Traveling information, Weather Nearest gas/charging station	Online news
	Black box systems (crash monitoring – Sep. 1st 2014)	Back2car	
Social Media	Facebook		newsfeeds, twitter

PRIVACY PROTECTION POLOCIES

This section will introduce the term privacy protection and provide the background of common protection policies that are mostly addressed by the European and in some cases by the U.S. law.

Privacy definitions

The general definition of privacy describes “The state or condition of being free from being observed or disturbed by other people” (Oxford, 2013).In particular data privacy can be defined as “The relationship between collection and dissemination of data ...” (Ethics point, 2013). Further privacy concerns “…exist wherever personally identifiable information is collected and stored – in digital form or otherwise” (Ethics point, 2013). A trivial form of personally identifiable information in the automotive context can be vehicle identification numbers (VINs), or more complex, location information (GPS information) allowing for traceability and making IDs negligible (see *Chapter 8*).

In the following section discusses the contrast between European and U.S. regarding privacy protection.

U.S. vs. European governmental privacy policies

A fundamental difference between the two legislations represents the treatment of sensitive personal data. In its basis the U.S. law does not provide one comprehensive statute conducting data protection or privacy issues but has a number of laws and

executive orders instead. It's mainly the Privacy Act of 1974 and the Computer Matching and Privacy Act offering laws that solely deal with personal data held by the federal government but however have no authority over the collection and use of personal data held by non-government parties. Further, the Computer matching and Privacy Protection Act adds regulations controlling the usage of computer matching. The term matching in this context refers to computerized comparison of individual data that shall determine the eligibility for Federal benefit programs (e.g. recouping payments, delinquent debts).Additionally the Computer Security Act assures the security of personally identifiable data in federal computer systems. Supplementary, there have been laws created by the U.S. legislation that are in a wider context related to privacy and data protection. They cover aspects like prohibiting the use of personally identifiable data from the census, protecting against disclosure of personal data gathered by the National Centers for Health Service and Research, revising the confidentiality and dissemination practices, making tax return information confidential and eventually having criminal penalties for illegal disclosures. A second type of law creating in the U.S. can be described as a responsive approach. It is that laws are created in reaction to observed abuses. One response was a restriction for the federal government to access records held by other sources. Until the era of the Internet, misuse of personal data held by public or private entities was not conceived by policymakers as a threat to privacy or personal liberty (Stratford et al., 1998).

In the starting process of adapting to the connected digital world, the European Union (EU) has had a predominant role regarding international decisions on information privacy. The dominance of the EU has been strengthened by the authority of EU Member Nations, as they coherently block data transfers from their country to third party nations

(EPCD, 1995).The term “third party countries” is also referring to such nations as the U.S., which according to the EU is missing “adequate” privacy protections (Hustinx, 1999). The sectoral privacy law in the U.S. divides the responsibilities for public and private data categories. As a consequence, in case data cannot be assigned to any of the available categories, they might not be protected at all (Solove et al., 2011).

Over the years the U.S. sector-by-sector approach stands still in contrast to the EU’s so called omnibus legislation, which treats personal data regardless whether they are private or public sector related. The global reaction prooves the EU as highly influential, whereas the U.S. appears to be an outlier regarding data protection. The EU Data Protection Directive established mutual rules for data privacy among its member states and set a three year deadline to adopt compliant legislation (Regan, 1995).

Eventually the U.S. law understood the importance of privacy protection and enacted data protection laws. Further the Commerce Department of the U.S. Federal Trade Commission (FTC) is now a full member of EU privacy conferences (Bamberger et al., 2011).

Due to the still existing EU supremacy regarding data protection legislation, the protection measures in this thesis shall be based on the European standards, briefly introduced in the next section.

Golden Rules

The so called “Golden rules” (listed in *Chapter 6*) are mounted into the German, British and most other European countries federal laws (HM Government, 2008). The act focuses on the protection against misuse of personal data in terms of data processing. The definition can e.g. be found as an annex to section 9 in the German “Federal Data Protection Act”, which is related to technical and organizational measures and therefore fundamental policy for big companies like Volkswagen dealing with various amounts of data.

Later on these rules shall be used as state of the art privacy policies in order to evaluate the presented privacy concepts.

This section will introduce standard and to a greater extend later applied data and IT-security mechanisms.

Certificates

Digital Certificates

Digital certificates supplement electronic messages with the purpose of providing authentication measures. The term authentication refers to the procedure that a receiver of a message is able to verify that the sender is the true sender. In order to send an encrypted message, an entity needs to apply for a digital certificate from a Certificate Authority (CA). The CA issues an encrypted digital certificate consisting of a public key and various identification data. For protocol reasons the CA's own public key is made available to the public via e.g. the Internet (Fig.1).

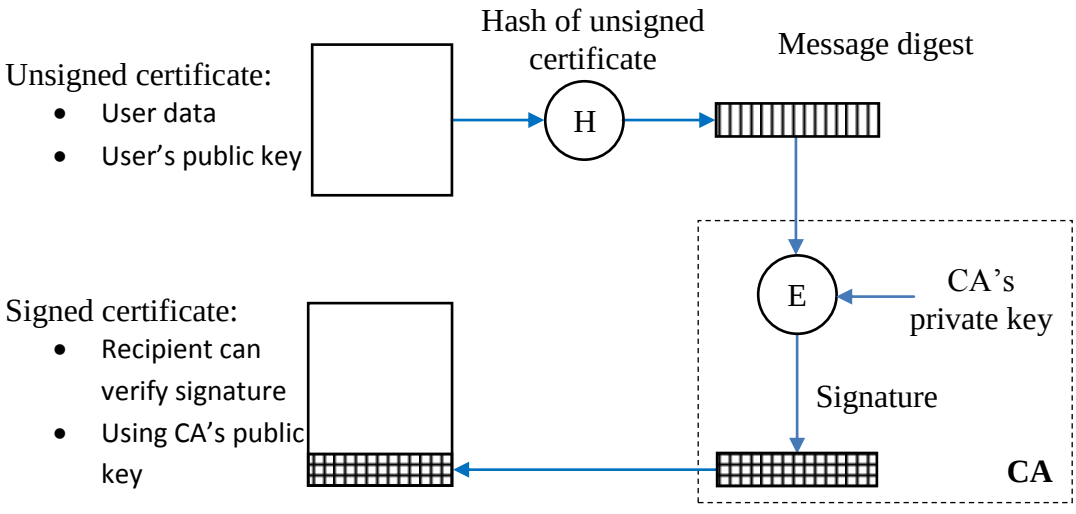

Figure 1: Digital signing (Code.google, 2013)

Now the recipient of the encrypted message makes use of the CA's public key in terms of decoding the digital certificate tied to the message. This action verifies that the certificate has been issued by the CA and also allows the receiver to obtain the sender's public key and identification data included in the certificate (Fig.2).

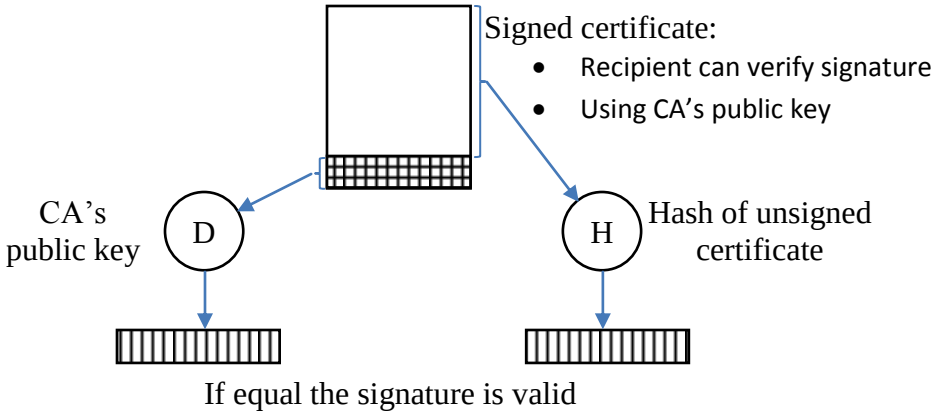

Figure 2: Digital verification (Code.google, 2013)

Based on this new knowledge the receiver is now able to send an encrypted response. The most commonly applied standard for digital certificates is X.509. Version 1 was introduced in 1988 integrated into the International Telecommunication Unit (ITU) X.500 Directory Services standard. Since then, two more revisions of the standard have been published including additional fields supporting directory access control, extensions for additional data regarding the certificate holder and the certificate usage. The main parameters of a certificate are listed in the following (Tab.6).

Table 6: Certificate parameters (RFC 5280, 2013)

Name	*Content*
Serial Number	Integer
Signature	Algorithm Identifier, Value: bit string
Issuer	Name
Validity	Not Before Time, Not After Time
Subject	Name
Subject Public Key Info	Algorithm Identifier, Value: bit string
Version	Integer {v1(0), v2(1), v3(2)}
Time	UTC Time, General Time
Unique Identifier	Bit string
Standard Attributes	Country
	Organization
	Locality
	Title
	Name
	Pseudonym

In general, the term X.509 denotes the latest, Version 3, unless the version number is specified differently (RFC 4158; 5280, 2013).

Certification Authority – CA infrastructure

The management of certificates is one of aspects covered by the commonly applied WAVE standard (Wireless Access in Vehicular Environment) in vehicle-to-x

(V2X) communications. The standard comprises information exchange among vehicles (private and public), confirmation with Certification Authorities (CA), which includes providing personal data in a wireless environment. Besides the CA the main parties involved are the car manufacturer (OEM) and the vehicle itself represented by the hardware security module (Fig.3).

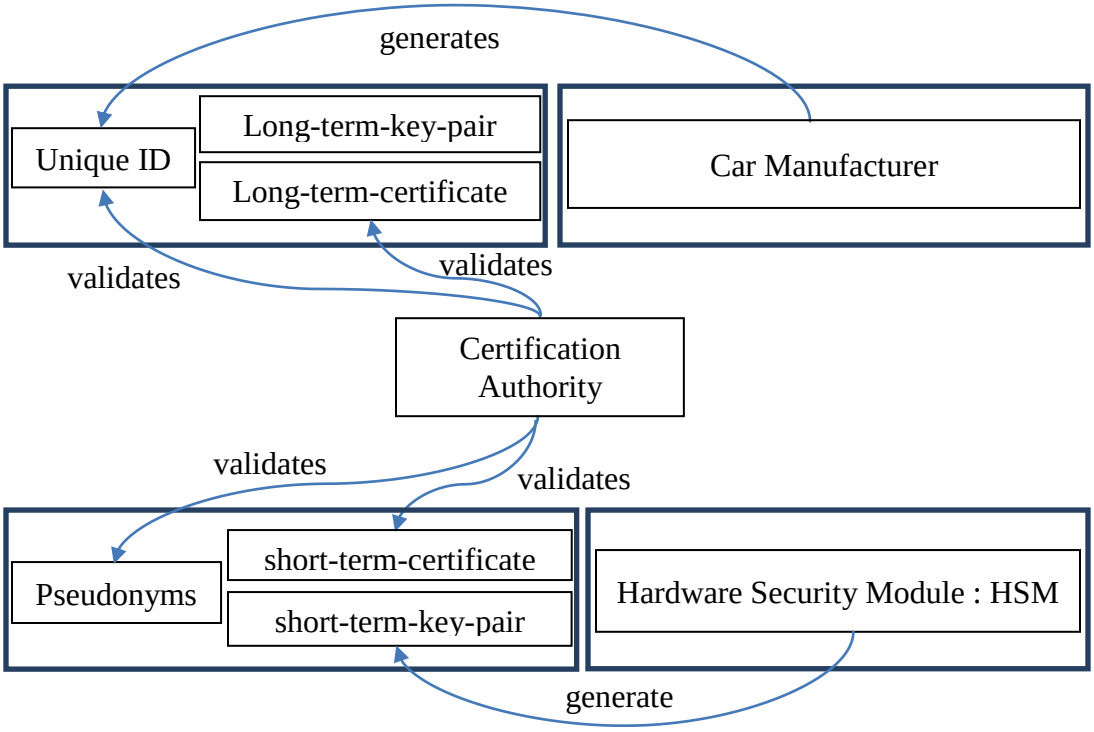

Figure 3: Certification management

The OEM generates a unique ID (usually the Vehicle Identification Number – VIN), a long term public-key pair linked to a long-term certificate. The certificate including the ID and other data shall then be validated by the CA. This process allows for an authenticated communication among the different entities. As a part of the in-vehicle communication architecture, the hardware security module (HSM) generates short term key pairs and sends them to the Certification Authority. The CA then validates them and generates related pseudonyms, both linked to a short term

certificate. There are different responsibility level for a CA, that is general regional limitations but also country limitations. In order to prevent identification based on certificates linked to certain regions or countries a mechanism called cross-certification has been implemented into the standard. In order to initiate the cross certification the vehicle with a certificate from region (A) would first have to authenticate itself to a CA responsible for region (B) by providing its existing long term certificate. Once a vehicle crosses the border to the new authority region, a new short term certificate, with a new set of associated pseudonyms and public key pairslinked to the new region, will be issued by the new CA. At the point, when the vehicle returns to its original region, the short-term certificate that was specifically generated for region B, will be revoked and the responsibility will handed back again to the original CA. Another important feature of certification revocation is used for authentication management. Vehicles or other communication participants like road side units (RSU) functioning as gateways within overall vehicular ad-hoc networks (VANETs) can be revoked in case of e.g. malicious behavior. The information about revoked participants is distributed in so called certification revocation lists (CRLs) that are broadcasted among the participants.

Cipher methodologies

Symmetric encryption

Symmetric cipher algorithms are based on so-called “secret-keys” shared between the communicating parties. The term “symmetric” refers to the fact that encryption and decryption are using the exact same key. The basic principle is illustrated in the graphic below (Fig.4).

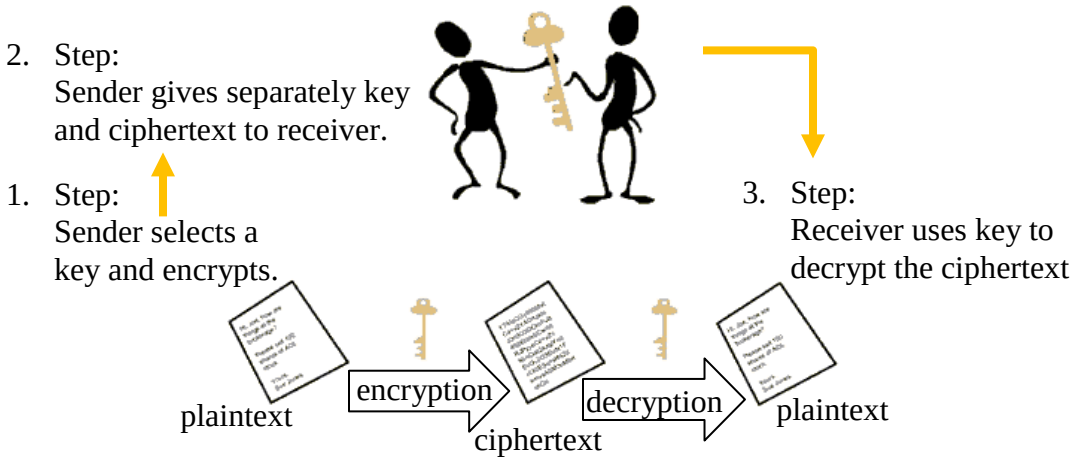

Figure 4: Symmetric cipher algorithm(MxRelease, 2013)

Data Encryption Standard (DES) was invented by IBM and the National Institute of Standards and Technology (NIST) in 1976, USA (see INV; FIPS, 1999). It belongs to the family of block ciphers and uses involution (a function is its own inverse) as main function and highly non-linear function as security mechanism. Complementarily it uses a key map providing keys for every round of involution. It takes two 32bit input blocks (L) and (R) as clear text and outputs 64 bit of cryptogram $\mathbf{Y} = (\mathbf{L'}||\mathbf{R'})$. The key map provides a different key (K) of 48bit length derived from a 64bit key for up to 16 rounds. The round structure was invented by the IMB engineer Horst Feistel, who migrated from Germany to the US in 1934. The already mentioned involution function is a self-inverting function (F) that applied twice compensates

itself. After all, nobody was able to break the involution mechanism since 1976 (Menezes, 2001).

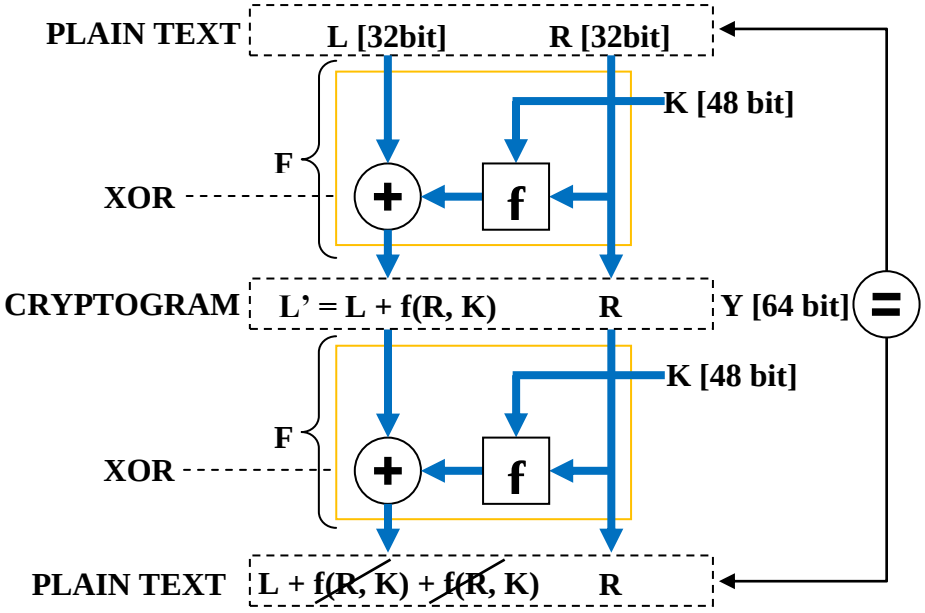

Figure 5: Data Encryption Standard (DES) with one key (K) encryption

In case there is more than one key involved a transposition phase will be added for (N) additional key blocks (Fig.6).

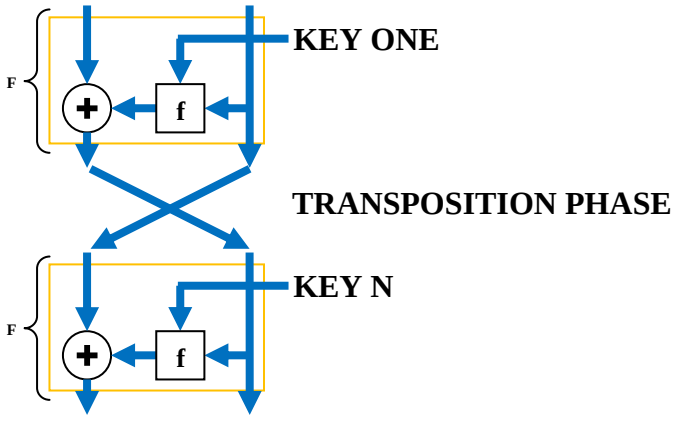

Figure 6: Data Encryption Standard (DES) with (N) key (K) encryption

An advancement of DES is the so called Triple DES (TDEA). It comprises three times cascaded DES, with three keys of each 56 bit length. TDEA offers a

comparatively simple method that increases the key size of DES to protect against such attacks. It comprises three times cascaded DES, with three keys of each 56 bit length. Accordingly, there are different keying options, with either all three keys being independent, only two keys independent, or all keys the same. The more keys are independent the longer becomes the total key length (FIPS, 2012). The longer the key length the stronger in fact becomes the cipher but also the computational complexity.

Asymmetric encryption

Asymmetric cipher algorithms are based on so-called "public-key" protocols. The term "asymmetric" is related to the fact that the encryption and decryption keys represent each other's inverse. Subsequently, when both keys come together they revoke the cipher and the secured message will appear. The public key of the receiver shall be used for encryption, whereas the sender's secret key is used as the signature. This guarantees that only the intended receiver can open and decrypt the message and allows the sender to authenticate it. The basic principle is illustrated in the graphic below (Fig.7).

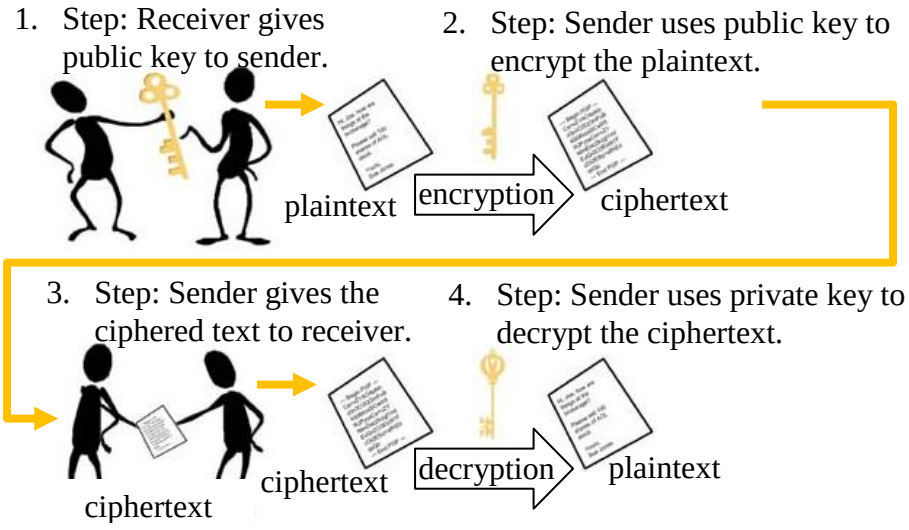

Figure 7: Asymmetric cipher algorithm (Data-Processing, 2013)

Digital Signatures

An additional application of public cipher algorithms besides encryption is to provide digital signatures as proof of authenticity of a digital message or document (Rivest et al., 1983). A common signature scheme is the RSA Digital Signature scheme, which shall be presented in the following reflecting the work of Rivest et al.

In order to have two parties communicating in a secure and authenticated manner, both need a public key and a public modulo(n). The latter shall be based on two secret preferably high primes generating a multiplicative inverse Group$\mathbb{Z}_n^*$.The highest order and the number of invertible elements in a multiplicative group can be determined with the Euler Function φ(n), with φ(φ(n)) as the number of units with the highest order (Delfs et al., 2007). Since RSA operates with exponents there are two related modulo defined in the scheme. According to Euler's Totient Theorem, that is$\boldsymbol{a^{\varphi(n)} = 1 \bmod n}$, such that (a) and (n) being relatively prime it holds that φ(n) is defined as modulo of the exponent, with $\boldsymbol{\varphi(n) = n - 1}$ for (n) being prime (Weisstein, 2013).

The parameter setup for the scheme is defined as follows (Rivest et al., 1983).

Public parameters

1. Public Keys
    a. For user A : E_A
    b. For User B: E_B
2. Public Modulo
    a. For user A: $N_A = p_A\, q_A$, defining the modulo for user A
    b. For user B: $N_B = p_B\, q_B$, defining the modulo for user B

Private parameters

1. Private Key user A and B: D_A, D_B
2. Secret large prime pairs of user A and B: (p_A, q_A) and (p_B, q_B)

The protocol comprises the following steps:

1. Step: User A generates cryptogram (Y), with E_B in modulo N_B, with

   $$\mathbf{Y = M^{E_B} \bmod N_B}$$

2. Step: User A generates signature (S), with proving content (C) and D_A in modulo N_A, with

   $$\mathbf{S = C^{D_A} \bmod N_A}$$

3. Step: User A sends (C, S) to user B
4. Step: User B decrypts cryptogram (Y) and verifies (S)
   a. $\mathbf{Y^{D_B} = (M^{E_B})^{D_B} \bmod N_B = M}$
   b. $\mathbf{S^{E_A} = (C^{D_A})^{E_A} \bmod N_A = C'}$
   c. $\boldsymbol{C \stackrel{?}{=} C'}$**,** proves authenticity of the sender

Blind Signature Schemes

The first introduction of blind signature schemes was published by Chaum, allowing messages signed by a third party without exposing any information about the message itself (Chaum, 1983, 1985). Blind signatures have various usability including anonymous access control, and digital cash.

In his work he extended the implementation of RSA signatures (Rivest et al., 1983) as follows. A client has a message m that needs to be signed by another party like e.g. a communication server, and the client does not want the server to know

anything about (m). Let (e, n) be the server's public key and (d, n) the private key, with n being the applied arithmetic modulo. The client generates a random unit (r) (an element that is invertible under multiplication is called unit; Giambruno, 2008). Another property of (r) is that it satisfies $\boldsymbol{gcd(r, n) = 1}$ stating that (r) is *relatively prime* to (n). In other words (r) and (n) have no common factor raise the level of protection (Johnston et al., 2009). The modular multiplicative inverse of (e) or (r) can be determined based on the Extended Euclidean algorithm (Koshy, 2007). An example design using the RSA Blind Signature shall give a more detailed insight on how the defined mechanism is designed (Goldwasser et al., 2008).

1. Step: *Server defines public directory and sends it to the client*
    a. Server defines public directory (e, n), with highly prime (e) as public key and public modulo $\mathbf{n = pq}$, with p and q highly prime and $\mathbf{gcd(e, n) = 1}$
    b. Sends public directory to client
2. Step: *Client computes blinded message and sends it to the server*
    a. Client defines message (m) and random unit (r)
    b. Client Computes blinded message $\mathbf{BM = (m\, r^{e})\ mod\ n}$
    c. Sends (BM) to the server
    - The server cannot derive any useful information from (BM)
3. Step: *Server signs blinded message and sends it back to the client*
    a. Signs the blinded message by computing

    $\mathbf{BMS = (BM)^{d}\ mod\ n}$ and sends it back to the client

4. Step: *Client extracts signed message*

    a. Client Computes $\mathbf{BS} = (\mathbf{BMS})\mathbf{r^{-1}} = (\mathbf{m}\,\mathbf{r^{e}})^{\mathbf{d}}\,\mathbf{r^{-1}}$, with

    $= \mathbf{m^{d}}\,\mathbf{r^{ed}}\,\mathbf{r^{-1}} = \mathbf{m^{d}}\,\mathbf{r^{0}} = \mathbf{m^{d}}$, that is

    by the server signed message (m) with private key (d)

    b. Obtains the true blind signature (BS) of (m)

The security of this signature scheme is implicit with the standard security argument that *factoring and root extraction remains computationally infeasible* (Gregg, J. A. et al., 2003). In general the signature scheme is unconditionally "blind" since (r) is chosen randomly and therefore does not allow the signer (here: the server) to learn about the message even if just mentioned computational infeasible problems can be solved.

Commitment Schemes

A commitment scheme describes a secret agreement or exchange of knowledge about certain information or message (Pedersen, 1991). In this thesis, the common Chaum-Pedersen commitment scheme will be applied and described in following. The message commitment scheme comprises two steps, the commitment and the opening *MC = (Commit, Open)*. The commitment executes *(c,sk)* ← *Commit(m,r)*, whereas the input message (m) and randomness (r) generate a commitment (c) to message (m) and secret key that is necessary to open the commitment. In the subsequent opening step the client executes *(m,r)* ← *Open(c,sk)*, whereas the inputs here commitment (c) and secret key (sk) generate the output message m and randomness (r) applied in the commitment.

Properties of a secure bit commitment scheme:

1. hiding: no knowledge about the message m is exposed
2. binding: committing to c and generating an opening (m', r') with m'≠ m is infeasible ((c) binds the client to message (m))

In other words, the commitment is hiding based on the fact that $\boldsymbol{h^r}$ is distributed uniformly over Group $\mathbb{G}$ and therefore hides $\boldsymbol{g^m}$ within the commitment $\boldsymbol{c = g^m h^r}$. The binding property is infeasible to break based on the assumption that the discrete logarithm over $\mathbb{G}$ for a polynomial-time adversary is infeasible (Harn, L., 1994).

An example design using the Chaum-Pedersen Commitment shall give a more detailed insight on how the defined mechanism is designed (Chaum et al., 1992).

System setup: The receiver chooses:

1. Group $\mathbb{G}$ of prime order p (so the discrete logarithm is hard to solve)
2. Generator g of order-q subgroup of $\mathbb{Z}_P^*$
3. Secret (a)
4. Scheme Parameter $\mathbf{h = g^a mod\ p}$
5. Commitment scheme $\mathbf{c = g^m\ h^r\ mod\ p}$

The sender chooses a particular message m ∈ $\mathbb{Z}_P$ and a random factor (r).

1. Step: *(c,sk)* ← *Commit(m,r)*
    a. Server defines public directory p, g, $\mathbf{h = g^a mod\ p}$, with (a) private
    b. Sends public directory to client
    c. Client computes commitment c'
    d. Client sends (c', m, r) to server

2. Step: *(m,r) ← Open(c,sk)*
    a. Server computes (c), with (h, g, m, r)
    b. Server verifies commitment with computed (c) and received (c'), that is $\mathbf{c'} \stackrel{?}{=} \mathbf{g^a h^r mod\ p}$

Zero Knowledge Proofs

So called Zero-Knowledge-Proofs-of-Knowledge (ZKPoK) were invented by Goldwasser, Micali and Racko in 1982 (Goldreich, 2002). The general approach of mathematical proofs is to provide all the necessary facts in order to prove that a statement is true. In contrast a ZKPoK does not reveal any facts. Zero-Knowledge Proofs (ZKPs) allow having another party prove that a statement is true. The other party will be completely convinced about the truth of the statement, but will not learn anything about it. In other words, the other party will gain zero knowledge (Barak, 2010). In this thesis, the Schnorr's ZKP will be applied suggesting a proof of knowledge for the discrete logarithm, which will be explained in the following.

Recollecting the setup from the previous parameters:

Commitment parameters

- Multiplicative group: $\mathbb{Z}_p$, with p being prime
- Message: m
- Commitment: c
- Randomness: r
- Secret key: a
- Generator: g
- Public parameter: $\boldsymbol{h} = \mathbf{g^a}$

Proof parameters:

- Additional randomness: $r_1,\ r_2,\ \gamma$

Identification protocol:

1. Step: Client computes (R_1, R_2)
   a. $\mathbf{R_1 = g^{r1}}$
   b. $\mathbf{R_2 = h^{r2} = (g^a)^{r2}}$
   c. Sends (R_1, R_2) to server
2. Step: Server chooses random γ
   a. With γ from $\mathbb{Z}_{37}$
   b. Sends γ to client
3. Step: Client computes (Z_1, Z_2)
   a. $\mathbf{Z_1 = r_1 + \gamma \times m \bmod \varphi(p)}$, with $\boldsymbol{\varphi(p) = p - 1}$
   b. $\mathbf{Z_2 = r_2 + \gamma \times r \bmod \varphi(p)}$
   c. Sends (Z_1, Z_2) to server
4. Step: Server verifies
   a. $\mathbf{(R_1 \times R_2) \times c^{\gamma} \stackrel{?}{=} g^{Z1} h^{Z2}}$

CHAPTER 3 – VEHICLE DATA EXTRACTION AND DISTRIBUTION FRAMEWORK

INTRODUCING THE FRAMEWORK

With respect to recent developments within the automotive industry, reliable range estimation for electric vehicles (EVs) has become a crucial feature when it comes to investing in this new technology. Online diagnosis would be based on the transfer of diagnostic data from the vehicle to the backend for immediate interpretation.

A more general approach is to combine various kinds of vehicle-data with GPS tags enabling completely new algorithmic opportunities. In this context, one goal of this architecture is to make the required vehicle-data (e.g., energy consumptions, current speed, and diagnosis data) available on backend for further processing.

Future developments could go into customized route recommendations with topics like navigated guidance in terms of route recommendations titled with e.g. “Scenic” vs. “Sport” routes. This approach will depend on existing driving style data combined with car model information and other useful criteria like the already mentioned GPS data.

The attempt of this first application concept is to construct an architecture that is able to collect live-data from on-board units inside the vehicle and send them back to the backend for further processing. Overall the approach of this service follows the

characteristics of connected applications that are making local data globally accessible. Further details on the data processing will be discussed in the section.

DATA PROCESSING

The data processing concept essentially comprises of four different phases, data extraction, data collection, data publishing and subscription (Fig.8).

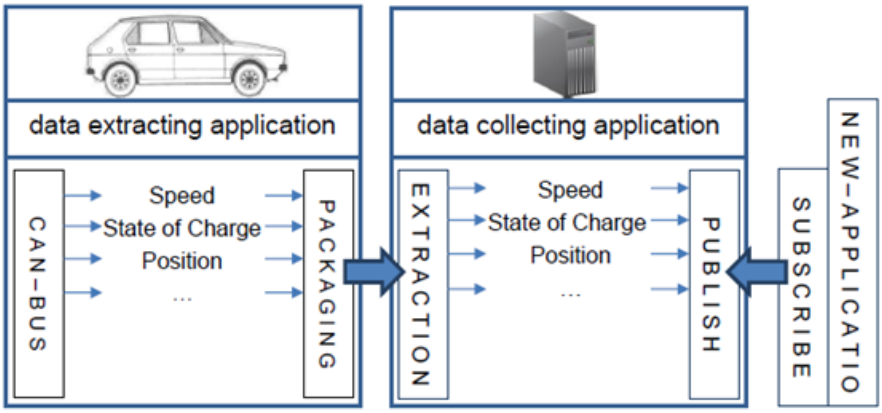

Figure 8: Data processing

The data extraction determines the relevant BUS data from inside the vehicle. This includes the generation of update packages and sending them to the backend. The data collection represents the extraction and processing of all the data belonging to one update package in order to make them available for an another independent application.

In the final step, the aforementioned independent application will be able to access the forwarded vehicle data by subscribing to context relevant data. The data packages consist of status information and currently measured values.

In general there are two approaches when it comes to connected applications which are related to either increasing or decreasing the data complexity. In the following sections both approaches will be specified with concrete use-cases.

POSSIBLE CONCEPT ALTERATIONS

This first approach represents an increased complexity level. It can result in applying complex data mining algorithms allowing for substantial long term statements. Regarding the attempt of achieving a more accurate and reliable range estimation it has been shown that an increase in data amounts can bring significant improvements (Ferreira, J.C. et al.). On the downside, data mining leads to various privacy issues and most attempts so far have used obscurity as a security mechanism. This reduces the probability of identifying a particular individual but still doesn't protect it appropriately (Cynthia Dworket al., 2010; Clifton, C., 2007).

Two alterations of the main concepts (profile generation and diagnosis statements) follow Each will be explained in their functionalities and privacy requirements.

One potential concept alteration could be the generation of driver profiles as an example for an increased data complexity. As already discussed, collecting data from the vehicle combined with certain logic and/or GPS data allows for various amounts of services, often referred to as connected services. For example, driving style matched with the car model can allow for route recommendations provided to the customer (Fig.9). Alternatively, it can help car manufacturers to improve their marketing

strategies by creating new vehicle configuration packages specifically for the discovered customer requirements. Considering electric vehicles, energy consumption over time enables analysis potential of the vehicle's charging behavior based on the driven routes. This way the customers can adjust their driving behavior to optimize their charging strategy. All these various statements shall be then integrated in the customer's personal profile.

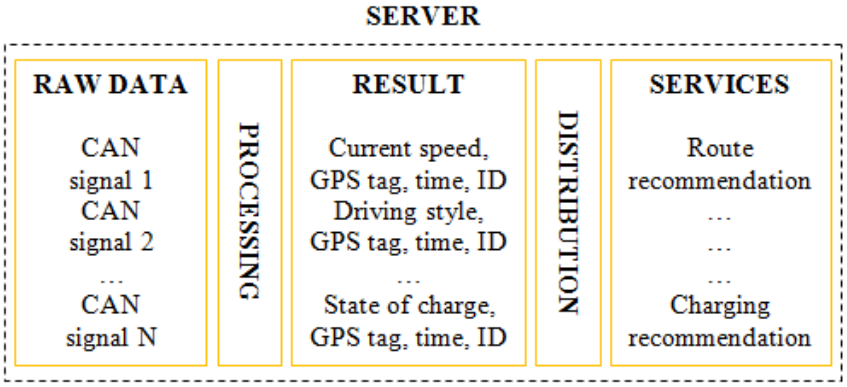

Figure 9: Protocol data processing concept

Another alteration can be to provide complex diagnosis statements to clients. They shall either provide immediate online maintenance (software updates, timing optimizations). In case of more complex diagnosis results, an automated appointment shall be made, based on information like vehicle type, driven miles, the actual diagnosis statement.

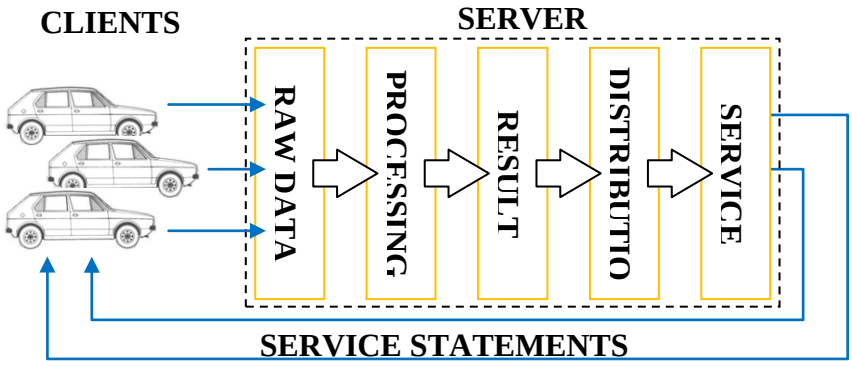

Figure 10: Protocol architecture

One can tell from the graphic above (Fig.10) that although the concepts have different applications, the overall communication architecture is quite similar. Accordingly, the next section will present a privacy management approach that shall be representative for the main and its related concepts.

CHAPTER 4 – PRIVACY PROTECTION INTEGRATION

HOW PRIVACY PROTECTION MECHANISMS VARY WITH THE APPLICATION

As a descriptive entry, it is helpful to briefly recollect the biggest challenges with connected applications these days. Data collections have become goldmines for a lot of different business models, e.g. customer profiles, long-term or historic data used by so called data mining companies for further analysis. Positive results for the end customer shall be more personalized services. The down-side of this well-intended approach, besides money, is that the traded good is personal and sometimes very private information (Fischermannet al., 2013; Facebook RepPortal, 2012, IMFSurvey Magazine, 2012).

One of the biggest wins from the connected world is the opportunity of nearly unlimited exchange of information (IMFSurvey Magazine, 2012) which in turn leads to the attempt of this use-case: achieving privacy protection while still being connected. In other words, the goal is to offer similar or even the same benefits to the customer but without exposing any sensitive, private or personal information.

In the process of finding solutions protecting personal data while staying connected this use-case and its presented alterations will show-case what data privacy requirements are necessary and where protection constraints begin. It has been shown that increasing data protection is not sufficient through merely making storage more

secure or enhancing access controls in order to solve all privacy concerns, due to the human factor (NYMITY, 2012).

The main goal with the presented framework is to make existing local data globally available. In the next step privacy concerns shall be addressed through the definition of privacy requirements.

PRIVACY PROTECTION CONCEPT FOR THE PRESETNED FRAMEWORK

The privacy concept discussed in this section shall be designed from the customer or client perspective. Since every service is based up on data that can be related to a private individual, it is important to protect it.

In the following, the necessary privacy requirements and its related mechanisms will be discussed for the “Live vehicle data extraction and distribution” concept.

Privacy requirements from the client perspective:

1. The client needs to give permission to share the data that is being processed by the available services
2. The raw data linked to the ID and location data must be protected
3. The processed plain text must not be accessible for unauthorized personnel
4. Optional. Additional trust bound between client and automotive company

Accordingly to the concept's requirements respective privacy protection mechanisms shall be explained and illustrated in example designs.

Requirement one can be satisfied with the following mechanism. This mechanism is fundamental for every system that claims to protect privacy. The setup of a privacy profile based on personal preferences and general policies given by the law should allow the client to manage what data are made available to the various service applications and which is kept private. IBM published on the fourth International Conference on Electronic Commerce Research (Bohrer, 2001) a technical approach for personal information and distribution. The communication matches is based on XML standards which matches perfectly with the automotive environment standards. Most telematics applications deal with the collection of information (in-vehicle measurements and status signals) that are usually streamed to the backend server in XML format (ASA, 2008; Telematics Update, 2013). Additionally, those data are mapped with very precise GPS data generating a very detailed personal picture of the driver. The necessity of these applications is often claimed by insurance companies since they can profit from the generated knowledge (ASA, 2008; Cognizant, 2012). It is obvious that there are cases (e.g. emergency cases) where GPS data can be very helpful in sending out assistance to a location provided by the individual in need. On the contrary, in a lot of other cases these GPS links may be utilized inappropriately by tracking individuals (ASA, 2008).

The IBM Privacy Services (IPS) system therefore provides several primary components based on IBM's Enterprise Privacy Architecture (EPA), handling privacy concerns for automotive telematics applications. The system includes automatic and

manual authorization for release of private data, matching the individual's and general law related privacy policies with those of data-requesters/application (see *Chapter 8* for details).

The following negotiation example is based on a dealer profile following the concept given by Bohrer et al. but translated into the automotive context.

1. A customer sets up an online appointment with a dealer for new car. He adds the note that he also wants to sell his old car.
2. The dealer receives the appointment request and asks in return for the name, address, salary and assets along with the privacy policy that the data will be used to approve credit. Additionally data about the old car will be requested like i.e. car make, mileage and year along with the privacy policy that these data will be used to determine the value of the old car.
3. The profile denies the first part of the request and offers to send an alias profile including salary and an asset range. On the other hand the request about the vehicle data will be accepted.
4. The dealer will accept the data but indicates that subsequently a final credit approval is not possible but only an analysis for a possible credit.
5. The customer sends the data for the credit analysis only as well as the vehicle data to determine the value of the old car.

Once the client has defined certain policies which shall be accompanied by common privacy laws, only authentic data requests that fulfill these policies shall be

permitted to the server (service provider).The authentication of the service request is based on two aspects, whether the requester is listed within the individual's privacy profile and whether the requester has the permission to view the data like e.g. sensor data. As another privacy protection mechanism, the possibility of "data mashup", autonomous data collection and integration through communicating applications must be prohibited (Soylu et al., 2012). Another more general goal of this approach is to build up trust between the client and the service application. Transparency in this context shall be utilized to establish this fundamental trust helping the client to understand what data are requested and what the purpose of this request is.

Essential for this approach is a classification of data in order to define handling rules respectively. Hence, a data classification shall be the result of this thesis evaluation.

Requirement two can be satisfied with the following mechanism. It needs to guarantee that all data linking to personal information, like IDs or location information will be protected. This kind of data, according to the main concept and its alterations, can be determined as the live data that is extracted from the vehicle's internal communication. In order to protect it, a secure communication path between the vehicle, the server and a secure storage needs to be established. Protecting the communication path shall prevent public attacks and the secure storage shall reduce the human factor by limiting the accessibility inside the company's backend.

The protection of the communication path can be achieved with either symmetric or asymmetric ciphers. In order to maintain the performance of the system, the symmetric approach shall be preferred at this point. A well-known cipher algorithm is

the Data Encryption Standard (DES, see Chapter 2). The initial DES cipher's key length of 56 bits was sufficient was sufficient in most cases, but due to increasing computational power, brute-force attacks had become feasible. Triple DES (TDEA) offers a comparatively simple method that increases the key size of DES to protect against such attacks. Since TDEA is mostly based on the DES algorithm, in the following an overview of the DES block cipher protocol shall be provided.

Upfront server and client need agree on a secret function (f) and secret keys (Ki). According to the protocol (L) and (R) and comprise a data block of 64 bit in total split up in 32 bit each. In the following a simplified example design will be presented showing the functionality of the algorithm.

1. Step: Agreement
    a. Server and client agree on:

       $f = R \times K_i^2$, with f extracting the $n = 4$ LSB of the actual result,

       $L'_i = L + f(R, K_i)$, with $(+)$ representing an XOR operation,

       and $K_i = \{2,3\}$
2. Step: Client encrypts data block: $(1011\ 1001)_2$ and sends it to the server
    a. Client assigns data blocks,

       with $L = 11 = (1011)_2, R = 9 = (1001)_2$
    b. Round 1: Computes $L'_i = L + f(R, K_1)$,

       with $f = (2^2 \times 9) = 36 = (100100)_2$

       extracting 4 LSBs $\rightarrow f = (0100)_2 = 4$

       $L' = (1011)_2 + (0100)_2 = (1111)_2 = 15$
    c. Transposition: $L = R = 9 = (1001)_2$,

and $R = L' = 15 = (1111)_2$

d. Round 2: Computes $L''_i = L' + f(R, K_2)$

$f = (3^2 \times 15) = 135 = (10000111)_2$

extracting 4 LSBs $\rightarrow$ $L' = (0111)_2$

$L'' = (1001)_2 + (0111)_2 = (1110)_2 = 14$

e. Transposition: $L = R = 15 = (1111)_2$,

and $R = L'' = 14 = (1110)_2$

f. Sends cryptogram $Y = (1111\ 1110)_2$ to server

Once the server receives the cryptogram it can be stored in the database and linked to e.g. a unique session ID for internal traceability but associated with a flexible expiration date. After the session expires the ID as well as the data shall be deleted. In order to accomplish the following steps, it is important that the processing of the decrypted data follows the policies determined by the Golden Rules (see *Chapter 2 and 6*).

3. Step: Server decrypts data block

a. Transposition: $L = R = 14 = (1110)_{\mathbf{2}}$

and $R = L'' = 15 = (1111)_{\mathbf{2}}$

b. Server assigns data blocks,

with $L = 14 = (1110)_2$, $R = 15 = (1111)_2$

c. Round 1: Computes $L' = L + f(R, K_2)$,

with $f = (R \times K_{i=2}) = (23 \times 15) = 135 = (10000111)_2$

extracting 4 LSBs $\rightarrow$ $f = (0111)_2 = 7$

$L' = (1110)_2 + (0111)_2 = (1001)_2 = 9$

d. Transposition: $L = R = 15 = (1111)_2$

and $R = L' = 9 = (1001)_2$

e. Round 2: Computes $L'' = L' + f(R, K_2)$

$f = (2^2 \times 9) = 36 = (100100)_2$

extracting 4 LSBs $\rightarrow$ $L' = (0100)_2$

$L'' = (0100)_2 + (0100)_2 = (1111)_2 = 11$

f. Receives data block $(1011\ 1001)_2$

What this means for the prior defined architecture is a security extension for parts of the components (Fig.11).

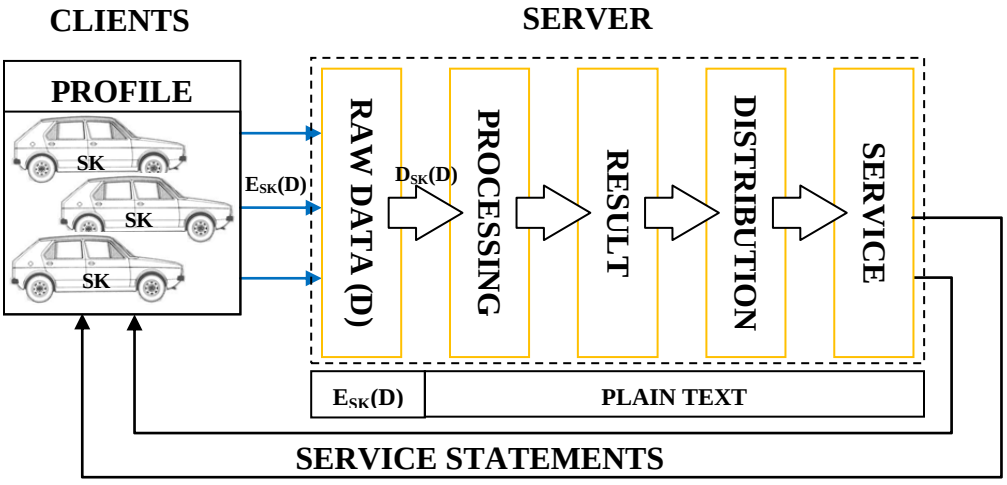

Figure 11: Privacy enhanced framework

Data are now communicated and stored encrypted according to the agreed secret key algorithm (secret key (SK), encrypted data $E_{SK}(D)$). Only requested processing data will be decrypted $D_{SK}(D)$ and made available as plain text. The secret key shall be only known by the automotive company and the protected onboard unit of the vehicle. Every vehicle's onboard unit is now equipped with a personal privacy profile such that only permitted data leaves the individual's vehicle. One part of the in Chapter 2

mentioned Golden Rules dealing with data protection is covered by the introduced mechanisms. It is assumed that the respective car manufacturer (OEM – Original Equipment Owner) has an established firewall system and secured data storage and therefore will be not part of this concept.

The plain text processing section of this concept (Fig.11) asks for access control policies mainly addressed by the Golden Rules and therefore already common knowledge in every automotive company's IT infrastructure. The assessment of these policies will be covered within evaluation later on in this thesis.

The optional trust bound between client and automotive company will only be mentioned but not integrated into the concept since the required mechanisms are not immediate privacy protecting mechanisms. In some cases trust mechanisms are even counterintuitive. They often expose additional personal data in order to identify the communicating parties or simply increase the overhead while the added value is questionable. A trust enhancing mechanism from the OEM perspective can be Digital Signatures that besides the actual encrypted message sends a signed piece of information or the actual message (the already encrypted message) that shall validate the party's identity. A signature is defined as the public of public key protocol like e.g. RSA (see Chapter 2). This will increase the overhead but in a manageable way, since the computational power is needed to verify a signature, which is fairly simple to do with the computational power on the backend side.

In order to enhance the trust between both parties, the usage of certificates (Chapter 2) allows authenticating the identity in both ways but also brings more computational overhead on both sides. It is important to mention though that in order

to keep privacy protected, only make use of pseudonymous public key certificates as they do not contain any identifiable information. Consequently they cannot be used to link to a specific client or to another pseudonymous certificate. Since trust enhancements are necessary, but not immediate privacy protection mechanisms rather than security mechanisms, they were not considered as a part of the concept.

Now that data are available as clear text, defined processing algorithms shall transform the extracted vehicle data into applicable statements that in turn shall be made available for the provided connected services. These services as a final step shall provide valuable recommendations or suggestions to the client.

In the following, two possible approaches for a communication initiation and profile setups are described. The communication itself shall be based on unique expiring session ID and can be distinguished as active and passive from the client perspective.

Active communication initiation:

1. Client onboard unit sends service request
2. Onboard unit checks with privacy profile for permission
3. Onboard unit sends permitted data to server
4. Server computes valuable statements
5. Service extracts statements and forms results
6. Service sends results back to client as recommendation or suggestion

Passive communication initiation:

1. Client onboard unit detects anomaly in diagnosis data
2. Onboard unit checks with privacy profile for permission
3. Based on profile settings
    a. Onboard unit sends diagnosis data to server
    b. Onboard unit informs client with in-car signal about necessary service
4. In case of a) server interprets diagnosis data extended services offers back to the client (e.g. service appointment data according to the diagnosis or applies immediate software correction if possible)

The profile setups below (Tab. 7) illustrate how the client can decide, what data shall be made available to a specific service for subscription. As an example, a service like a *travel guide* shall be made available based on several levels of exposed *personal related data*. The chosen data in this example are *destination location*, *current location* and *driving style*. Driving style hereby can represent slow/fast drivers, driving with adaptive cruise control (ACC), etc.

Table 7: Profile variations for a travel guide service

Profile 1		***Service List 1***
Allow	Deny	

subscription to chosen destination locations		• Destination information
		  ○ Weather
		  ○ Current events
		  ○ Extended sight information

	subscription to current location	
	subscription to driving style	

Profile 2		***Service List 2***
Allow	Deny	

subscription to chosen destination locations
subscription to current location
	subscription to driving style	• Destination information
		  ○ Weather
		  ○ Current events
		  ○ Extended sight information
		• General route recommendation
		  ○ Scenic routes
		  ○ Charging/refuel recommendations

Profile 3		***Service List 2***
Allow	Deny	

subscription to chosen destination locations
subscription to current location
subscription to driving style
		• Destination information
		  ○ Weather
		  ○ Current events
		  ○ Extended sight information
		• General route recommendation
		  ○ Scenic routes
		  ○ Charging/refuel recommendations
		• Customized routes
		  ○ Relaxed routes
		  ○ Sport, curvy routes

There are two important aspects to be noted at this point. On the one hand, the various profile setups shall demonstrate *flexibility* regarding the exposed data, based on the client's decision. On the other hand, it shows the subsequent service limitations according to the client's decisions.

In the following the hotspot identification protocol as an already existing concept will be explained with its functionalities and privacy requirements (Raghunathan et al., 2012).

PRIVACY PROTECTION CONCEPT FOR THE HOTSPOT IDENTIFACION PROTOCOL

Protocol overview

One approach for decreasing the data complexity is to reduce the amount of combined statements for simple interpretations. A more extreme approach regarding decreasing the amount of complexity can be found by looking at voting protocols. These protocols are dealing with rather trivial but effective statements by sharing quantity statements which as a consequence have no need for exchanging personal data. Having clients exchanging information through a distribution server makes it possible to filter location information so that individual identities are kept privately (Raghunathan et al., 2012).

The fundamental idea of this approach is based on a distribution server relaying between the clients. This way, identifying information (like e.g., location information or identities) can be filtered so that each individual's privacy is protected. Statements based on quantities derived from location based occurrences, like a lot of people in one spot, a lot of pictures taken in one spot, a lot of cars of one model or make in the same area can allow for hotspot identification. A lot of people in one spot can identify big events, a lot of pictures in one spot can identify points of interests (POIs) and a lot of cars of the same model or make can be an indicator for an optimal workshop or dealer position.

The architecture is divided into two communication protocols. The first is designed as a registration protocol based on an authenticated channel (Fig.12).

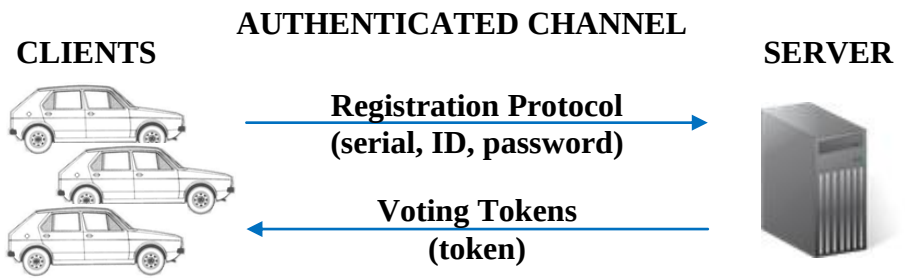

Figure 12: Hotspot identification registration protocol

Each client or automotive customer registers with an ongoing serial number to expand the uniqueness entropy, branded with a time-stamp for validation purposes and a corresponding ID and password. The reason for the registration phase is to equip each client with a voting token that will allow for participation in several votes per defined time period.

As already mentioned, the second is designed as voting protocol based on an anonymous channel (Fig.13).

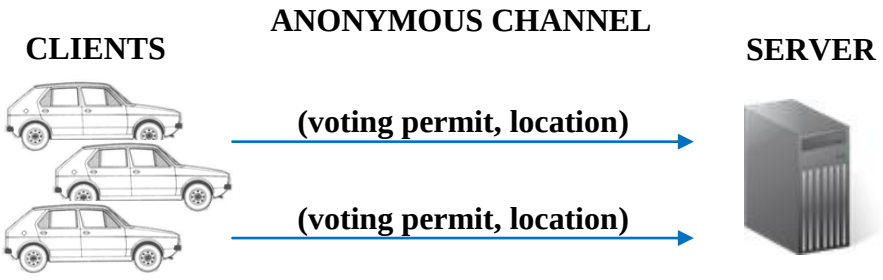

Figure 13: Hotspot identification voting protocol

After every voting period the client would have to register again to renew the voting permit. The re-vote feature is to guarantee up-to-date votes. A client's vote in this protocol provides a location statement including a position and auxiliary data like further comments.

The server, on the other hand, validates each vote and then only learns the location, not who is at that particular location. In the next step the server defines so-called rounded hotspots, hotspot areas, wherein each vote increments a vote count whose final tally will be published to all clients in the end. The total count is a quantity statement and tells each client about current hotspots.

The defined privacy requirements and its related mechanisms that shall protect each client's privacy will be discussed in the next section.

Identified privacy protecting mechanisms in the protocol

The focus of privacy requirements shall be based again on the customer or client perspective.

Privacy requirements from the client perspective:

1. Password and ID must be protected
2. Client ID must not be linked to submitted vote
3. Signatures must not link to the client's ID
4. The precise location must not be revealed to the server

Accordingly the concept's requirements respective privacy protection mechanisms shall be explained and illustrated in example designs.

Requirement one can be satisfied with the following mechanism. It needs to allow for a secret agreement and exchange of knowledge about the client's ID and password in order to guarantee authenticity (Pedersen, 1991). Common mechanisms to hide an agreement or commitment are so-called commitment schemes (see Chapter 2). In the

following an example design will be presented showing the functionality of the scheme.

The sender is this scenario will be represented by client and the receiver by the server. The client chooses a particular message $m \in \mathbb{Z}_P$ and a random r, with m representing the bit code of the defined password and ID password and ID. For computational ease a decimal setup in $\mathbb{Z}_{37}$ will be used.

1. Step: $(c, sk) \leftarrow \text{Commit}(m = 6,\ r = 3)$
    a. Server defines public directory $p = 37,\ g = 2,\ a = 5,$ and $h = 2^5 \bmod 37 = 32$
    b. Sends public directory to client
    c. Client computes commitment $c = 2^6 \times (2^5)^3 \bmod 37 = 27 \times 23 \bmod 37 = 29$
    d. Client sends $(c = 29,\ m = 6, r = 3)$ to server
2. Step: $(m, r) \leftarrow \text{Open}(c = 29,\ sk = r = 3)$
    a. Server computes (c), with $h = 2^5,\ g = 2,\ m = 6,\ r = 3$
    b. Server verifies commitment with computed c and received c, that is $29 \stackrel{?}{=} 2^6 \times (2^5)^3 \bmod 37 = 221 \bmod 37 \equiv 29$

Now server and client have agreed to hidden message (m) with a binding commitment (c). As a consequence, only the client with the correct ID and password will be eligible to vote. The client is now responsible to protect both from the public in order to avoid misuse like manipulating entire voting phases.

Requirement two can be satisfied with the following mechanism. So-called Zero-Knowledge-Proofs-of-Knowledge (ZKPoK) allows having a third party prove that a statement is true. This other party will be completely convinced about the truth of the statement, but will not learn anything about it. In other words, the third party will gain zero knowledge (Barak, 2010).In this thesis the Schnorr's ZKP will be applied suggesting a proof of knowledge for the discrete logarithm, which will be explained in the following with a number example.

This mechanism shall be used during the voting phase preventing the server from being able to link a vote to a registered client. This shall be achieved by having the client proving to the server that he knows his ID and password without revealing it.

Recollecting the setup from the previous parameters:

- Commitment parameters
    - Multiplicative group: $\mathbb{Z}_p$, with $p = 37$ being prime
    - Message: $m = 6$
    - Commitment: $c = 2^{21} = 29$
    - Randomness: $r = 3$
    - Secret key: $a = 5$
    - Generator: $g = 2$
    - Public parameter: $h = g^a = 2^5$
- Proof parameters: randomness: $r_1 = 7,\ r_2 = 11,\ \gamma = 13$

Identification protocol:

1. Step: Client computes (R_1, R_2)
    a. $R_1 = g^{r1} = 2^7$
    b. $R_2 = h^{r2} = (g^a)^{r2} = (2^5)^{11} = 2^{55}$
    c. Sends (R_1, R_2) to server
2. Step: Server chooses random γ
    a. With γ from $\mathbb{Z}_{37}$
    b. Sends γ to client
3. Step: Client computes (Z_1, Z_2)
    a. $Z_1 = r_1 + \gamma \times m \bmod 37$

       $= 7 + 13 \times 6 \bmod 36 = 85 \bmod 36 = 13$
    b. $Z_2 = r_2 + \gamma \times r \bmod 36$

       $= 11 + 13 \times 3 \bmod 36 = 50 \bmod 36 = 14$
    c. Sends (Z_1, Z_2) to server
4. Step: Server verifies
    a. $(R_1 \times R_2) \times c^{\gamma} \stackrel{?}{=} g^{Z1} h^{Z2}$

       $(R_1 \times R_2) \times (g^m h^r)^{\gamma} \stackrel{?}{=} g^{Z1} h^{Z2}$

       $2^7 \times 2^{55} \times (2^{21})^{13} \stackrel{?}{=} 2^{13} \times 2^{70}$

       $2^{335 \bmod 36} \stackrel{?}{=} 2^{83 \bmod 36}$

       $2^{11} \equiv 2^{11}$

Now Server has proof of the fact that the voting client has knowledge about his ID and password and therefore is a registered participant. At the same time the server did not learn anything about the ID and password so there is no chance that he can link

the collected hotspot vote to a specific client. In other words location privacy is provided.

Requirement three can be satisfied with the following mechanism. So-called blind signature schemes (BSS) allow for approving certain content achieved by a third party without exposing any information about the content itself (see Chapter 2). The content here is represented through the client's ID and password during the registration and voting phase and therefore must not be exposed to the server in order to prevent any linking to the client's ID.

In the following an example design will be presented showing the functionality of the scheme in $\mathbb{Z}_{161}$.

1. Step: *Server defines public directory and sends it to the client*
    a. Server defines public directory (n, e), with $(e = 35)$ as public key and public modulo $n = 7 \times 23 = 161$,

       and $\varphi(161) = (7-1) \times (23-1) = 132$

       and $\varphi(132) = 23 \times 3 \times 11 \times \left(1 - \frac{1}{2}\right) \times \left(1 - \frac{1}{3}\right) \times \left(1 - \frac{1}{16}\right) = 40$
    b. Computes secret key $d = e^{-1} = -49 = 83$ (in mod 132), (Tab.7) with (e) relatively prime to φ(n), as $\gcd(132,35) = 1$ is true.
    c. Sends public directory to client

Table 8: Multiplicative inverse of (d) in $\mathbb{Z}_{132}$

φ (n)	*e*	*B1*	*B2*	*q*	*r*
132	35	0	1	3	27
35	27	1	-3	1	8
27	8	-3	4	3	3
8	3	4	-15	2	2
3	2	-15	34	1	1
2	1	34	-49	2	0

2. Step: *Client computes blinded message and sends it to the server*
    a. Client defines message $(m = 2)$ and random unit $(r = 3)$
    b. Client Computes blinded message

    $BM = 2 \times 3^{35} \bmod 161 = 2 \times 124 = 87$ (in mod 161)

    c. Sends $(BM = 87)$ to the server

    ➢ The server cannot derive any useful information from (BM)
3. Step: *Server signs blinded message and sends it back to the client*
    a. Signs the blinded message by computing

    $\rightarrow BMS = 87^{83} \bmod 161 = 2^{83} \times (3^{35})^{83} \bmod 161$

    $\rightarrow = 2^{83} \times 3^{35 \times 83 \; mod \; \varphi(161)} \bmod 161$

    $\rightarrow = 2^{83} \times 3^{35 \times 83 \bmod 132} \bmod 161 = 2^{83} \times 3^{1}$

    $\rightarrow = 18 \times 3 \bmod 161 = 54$ (in mod 161)

    b. Sends (BMS = 54) back to client

4. Step: *Client extracts signed message*

    a. Client Computes

    $BS = 54 \times 3^{-1} = 54 \times 54 \bmod 161 = 54 \bmod 161 = 18,$ with $3^{-1} \bmod 161 = 54$, (Tab.8)

    b. Obtains the true blind signature to the message ($BS = 18, m = 2$),

    c. Prove: $m = (BS)^e: 2 \times 18^{35} \bmod 161 \equiv m = 2$

Table 9: Multiplicative inverse of (r) in $\mathbb{Z}_{161}$

n	***r***	***B1***	***B2***	***q***	***r***
161	3	0	1	53	2
3	2	1	-53	1	1
2	1	-53	54	2	0

Requirement four can be satisfied with the following straight-forward security proof. It shall be shown that location privacy requirement is satisfied and therefore the server has no knowledge about the exact location of the client.

The following assumptions and definitions are fundamental to show that location privacy is present:

It needs to be assumed that the client only transmits location and auxiliary data to the server, such that the auxiliary data as additional location information. It is required that the auxiliary data provided by the client shall not reveal any information about the client's ID. The definition IND-LP Raghunathan et al. are giving states that the ability to distinguish between two location data sets D_0 and D_1 must be negligible. For the

full system proof, the same indistinguishability needs to apply not only for the aggregated location information, but also for the auxiliary information.

Further Raghunathan et al. define two datasets D_0 and D_1 as *neighbors* if there exists two tuples in D_0(ID,loc) and (ID', loc') such that upon swapping only the ID's D_0 becomes D_1. Further they define *view (location information (D), private location hotspot protocol (PrivLHS))* as PrivLHS executed over location information D, returning the entire transcript of all interactions between the server and the client. As D_0 and $D_1 \in D$ are neighbors, both produce the same aggregated information. In order to satisfy the location privacy requirement, all neighboring datasets D_0, D_1, view(D_0,PrivLHS) and view(D_1,PrivLHS) must be (computationally) indistinguishable.

Informal proof. Raghunathan et al. go on and define intermediate datasets $I_0 = D_0$, $I_{1,\ldots,}I_{n-1}$, $I_n = D_1$, with I_j and I_{j+1} being neighbors with n denoting the number of clients. With the introduction of intermediate datasets and the prior defined *neighbor* definition, a neighbor relation can be identified between e.g. I_1 and $D_{0.}$ Given an increasing order of ID one can say that after swapping neighboring IDs e.g. the D_0 entry matches D_1. From the beginning of the proof, where I_j and I_{j+1} being defined as neighbors, it follows that transcript of all interactions view ($I_{j,}$PrivLHS) ≈ view (I_{j+1},PrivLHS). Subsequently the standard hybrid argument shows that view (D_0,PrivLHS) ≈ view (D_1,PrivLHS) which completes the proof of location privacy (the full proof - Raghunathan et al., 2012).

CHAPTER 5 – IMPLEMENTATION

This chapter shall first give an overview on the concept implementation including the major software components. Second, it will illustrate the data translation starting from the vehicle source passed on to the back-end for further processing. In order protect the intellectual property of the Volkswagen Group the following implementations shall only be described in its fundamentals.

IMPLEMENTING THE PRESENTED FRAMEWORK

This section will give an overview on the concept implementation including the major software components (Fig.14).

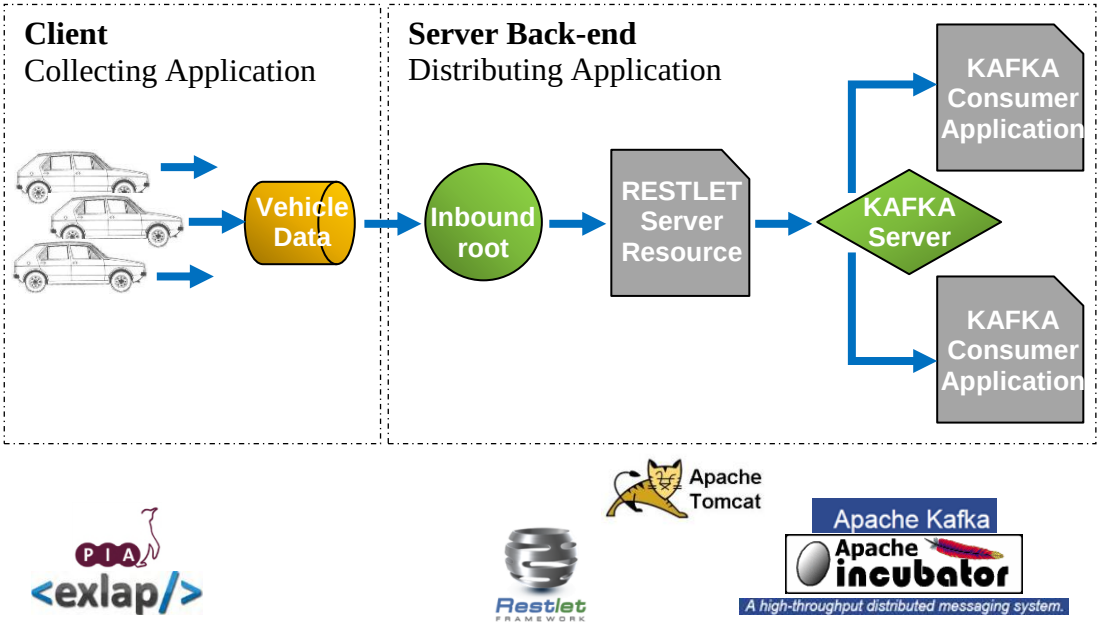

Figure 14: Software overview of the implemented framework

In the first step a Volkswagen proprietary software product called Exlap extracts the sensor and other data from the in-vehicle communication. Exlap then sends this data via HTTP to the back-end server. Exlap is a domain specific protocol to transport vehicle specific data (e.g. sensor information, state information) in a non-binary format over the "wire" to other devices and domains. The Exlap protocol suite also defines the specific bindings for different transport layers, namely Bluetooth, TCP/IP (WLAN) and representation via the HTTP protocol. Exlap is primarily aimed to provide a generic, uniform and universal access based on the requirements and restrictions of today's and tomorrows (mobile information) devices and their underlying platforms, e.g. Java/JavaScript in web browser environments. Exlap relies on the use of existing standards (i.e. XML, UTF-8 encoding) to not "reinvent the wheel" and encourage the simple processing of the transported data by leveraging the native XML processing capabilities of today's platforms (Fricke et al., 2009).

The following detail has been left out of the figure (Fig.14) above since it does not have any impact on the functionality of the architecture but shall be mentioned for the sake of completeness. In this implementation, as a temporary solution, the extracted live vehicle data in XML format is first sent to the Volkswagen back-end in Germany to be preprocessed before it is sent back to Volkswagen ERL back-end in Californian. The communication between Germany and California is based on HTTP.

For the implementation, the RESTlet architecture shall be used. REST (Representational State Transfer) is used for distributed systems such as the World Wide Web. It makes use of HTTP client connectors representing a software element that enables the communication between components (Fig.15). So called RESTful

architectures are based on client server relationships, such that a client sends a request to a server that will process the request and then send back a response (Fig.15).

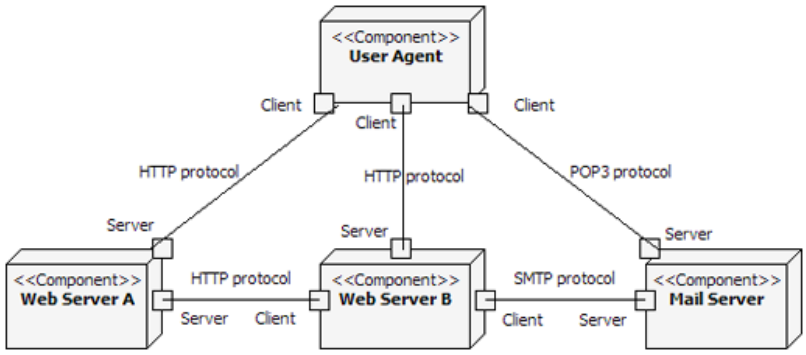

Figure 15: RESTlet client-server architecture (Restlet, 2013)

The smaller boxes shall represent the connector enabling the communication between components which are embodied by the larger boxes. The links denote the various communication protocols (HTTP, SMTP, etc.) that can be used. There are two very important characteristics. First, a client can have several resources to fulfill different tasks (Fig.15). Second, every resource can act as client or server, such that the data requested by one server can be made available to others (Fig.15). As the architecture name already indicates the communication is based representations of resources that are communicated between client and server. The representation of a resource is realized in the form of a document describing the current state of a resource (Restlet, 2013).

Once the current XML representation of the collected raw live data from the vehicle is received via HTTP POST on the ERL back-end side, it will be parsed into the slimmer JSON (JavaScript Object Notation) format. JSON is built on two structures, the collection of name/value pairs and an ordered list of values (Json, 2013). The JSON API represents a hierarchical structure (Fig.16). The two types

applying for this implementation are JSON Objects and JSON Primitives together describing a two dimensional relation. The available data are either represented by an immediate JSON primitive (name/value pairs, e.g. engineSpeed/0), or a JSON object with the name of a sensor group (e.g. heating) and a value representing several sensors in the form of name/value pairs (e.g. seatHeating/value, windowHeating/value, etc.).

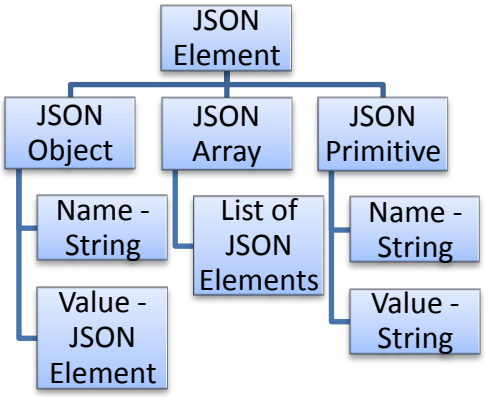

Figure 16: Hierarchal representation of the JSON API

After the explicit vehicle data are extracted they shall be fed into a Publish/Subscribe protocol, where data are made available for interested subscribers. This implementation uses the Kafka platform as messaging system that was initially developed at LinkedIn and is now used by multiple companies for all kinds of data pipeline and messaging (Kafka, 2013). In general, publish subscribe mechanisms introduce an extended communication infrastructure by i.e. adding topics and providing listening applications with subscribing capabilities (MSDN, 2013).

The task of publishing data in the form of a message that is linked to a topic is performed by a so-called producer. Consumers subscribe to a topic and accordingly receive every message that is published under it. The distribution of messages happens in such a way that each consumer process has its consumer group and each message is delivered to precisely one process within a group. This allows for two options. First,

various processes or machines can logically perform as a single consumer (queue semantic). On the other hand, in order to have every consumer receive the same published message, each of them needs to be in its own group (topic semantic). Eventually, Kafka has one more benefit regarding large data that regardless ofthe amount of consumers per topic, every message is stored just one time (Kakfa, 2013).

What this means physically is described by the following. There are three parties, the producer and consumer as already mentioned and the distribution server. On the server there are two components implemented, the Kafka-Server itself and the "Zookeeper." The Kafka-Server acts as broker (usually another machine) in that those messages are physically sent to a server acting as a broker. The broker is caching the data that are pushed from the producer and then pulled from the subscriber when ready to consume (Fig.17).

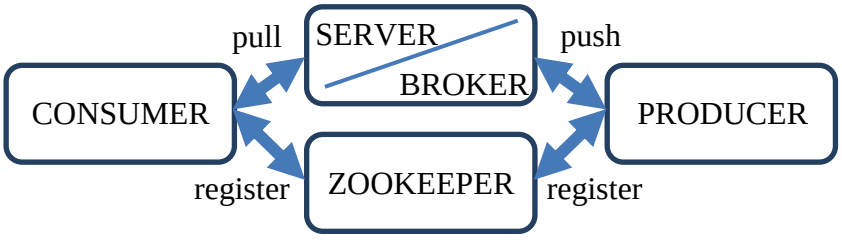

Figure 17: Physical communication in Kafka

Producer and consumer can be started dynamically anytime. The Zookeeper coordinates producer and consumer, in terms of meta-data registration (e.g. available topics, flow control, etc.) by each broker (Kafka, 2013).

SENDING SAMPLE IS SENT FROM THE VEHICLE TO THE BACKEND

This section will illustrate the data translation starting from the vehicle source passed on to the VW ERL back-end for further processing (Fig.18).

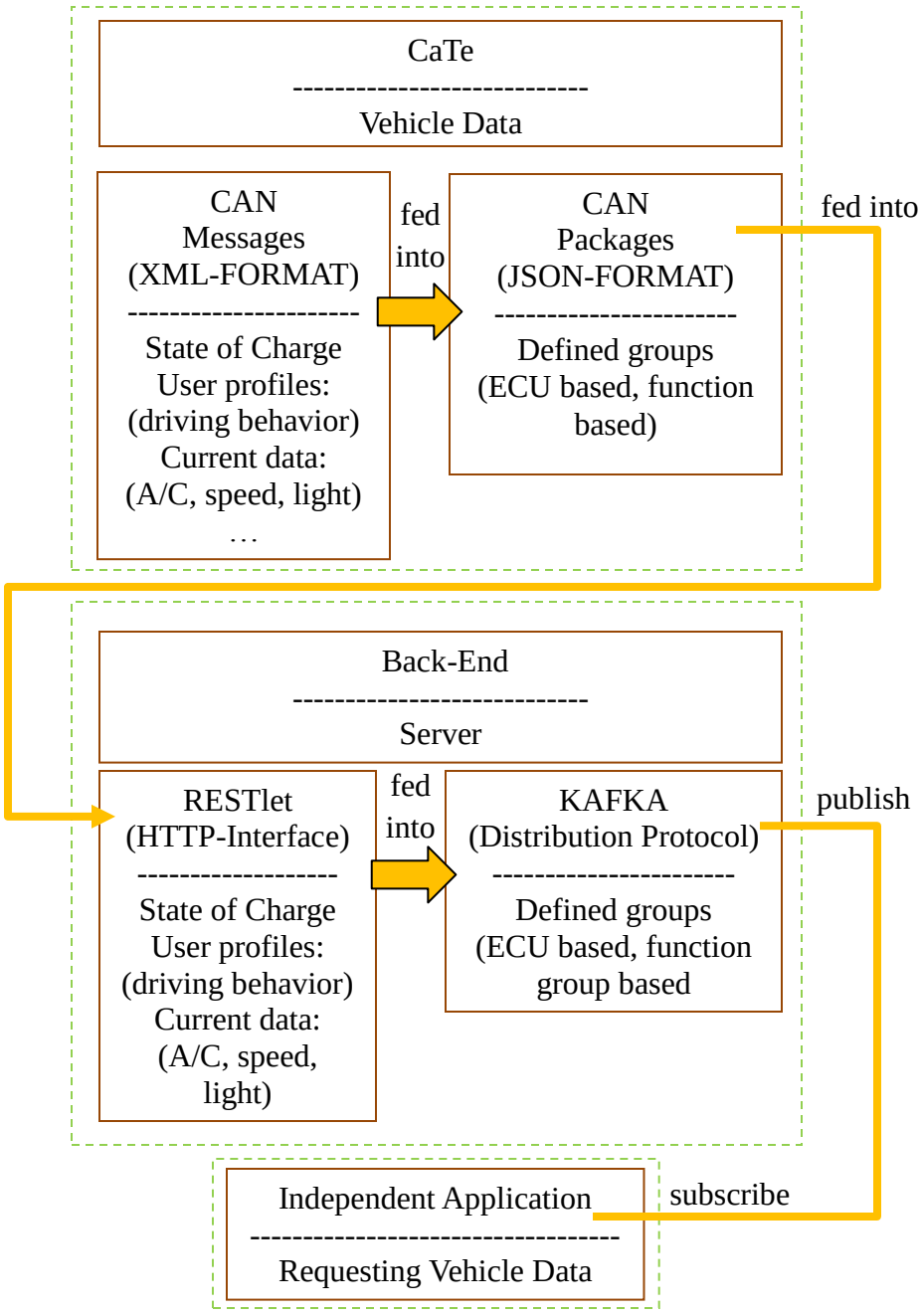

Figure 18: Data translation overview

According to the overview, the major parts of the implementation will now be presented and explained. The code is written in the language Java 1.7 using the following non-standard public APIs (Tab.9):

Table 10: API documentation overview

API-Name	***Documentation***
Org.json-20090211.jar	www.json.org
Org.restlet.ext.servlet.jar Org.restlet.jar	www.restlet.org
Scala-library-2.8.0.jar	http://www.scala-lang.org/
Gson-2.2.2.jar	http://code.google.com/p/google-gson/
Kafka-0.7.1.jar	http://kafka.apache.org/
Zkclient-0.1.jar	http://people.apache.org/~mmorel/apache-s4-0.5.0-incubating-doc/javadoc/
Zookeeper-3.4.0.jar	http://zookeeper.apache.org/
Log4j-1.2.15.jar	http://logging.apache.org/log4j/1.2/
Slf4j-1.7.0.jar	http://www.slf4j.org/

At first the recorded vehicle data needs to be translated into Java objects, which is mainly achieved by the following lines of code. The notation (…) indicates missing code representing internal setups and knowledge and therefore shall remain intellectual property of Volkswagen. Further class headers shall not be included in order to maintain readability.

This first class represents a Restlet client sending the extracted live-data to the back-end server.

```
public class ClientCall {

    /**
     *
     * Testing class Substitutes the Cargate Project in case there is no car
     * connection Simulates live-data
     *
     * @param args
     * @throws IOException
     * @throws JSONException
     */
    public static void main(String[] args) throws IOException, JSONException {

        // --- Preloads file containing logged live-data
        String xmlData = readfile("C:\\...\\erl_data.xml");

        JSONObject jObj = XML.toJSONObject(xmlData);
        String jsonString = jObj.toString();

        // --- Defining Vehicle Identification Numbers to be used as Kafka
        // GroupIDs
        String vin1 = "VWGEF9BP2CD054765";

        // --- Calling the Restlet Resource
        ClientResource resource = new ClientResource("http://.../hello");

        // --- Creating the Restlet data frame
        // as a Form consisting of key-value pairs
        Form form = new Form();
        form.add("VIN1", vin1);
        form.add("car1", jsonString);

        // --- Converting the data frame into a Web Representation(
        // --- so it can be send posted to the server
        Representation rep = form.getWebRepresentation();
        resource.post(rep);
    }
    (...)
    }
}
```

In the next step the Server Application class accepts the request and routes it to corresponding Server Resource class.

```
public class FirstStepsApplication extends Application {
    /*
     * Creates a root Restlet that will handle all incoming calls
     */

    @Override
    public synchronized Restlet createInboundRoot() {
        // Creates a router Restlet
        // that routes each call to a new instance of NewServer
        Router router = new Router(getContext());

        // Defines only one route
        router.attach("/hello", NewServer.class);

        return router;

    }
}
```

Now the vehicle data will be handed over to the Kafka Server class that functions as interface between the HTTP-Post resource and the Kafka server running the Kafka producer class

```
/**
 * Resource which has only one representation.
 */
public class NewServer extends ServerResource {

    // @ Representation: containing key-value pairs
    // VIN-update,
    @Post("String")
    public String posthandle(Representation entity) {

        Form form = new Form(entity);
        // Latest update that needs to be published
        String car1 = form.getFirstValue("car1");
        // Publisher containing the current update
        NewProducer producerThread = new NewProducer(car1, canList);
        producerThread.start();

        return "UpdateSent";
    }
}
```

The following producer class is responsible for the explicit vehicle data extraction that is available for each available sensor group. The data extraction is indicated as "Parser" and shall remain intellectual property of Volkswagen.

```
public class NewProducer extends Thread {
    // Needed instantiations
    // @kafka producer - needed to publish the update
    // @properties - needed for the communication setup
    // @gson - provides easy json string conversion
    private final kafka.javaapi.producer.Producer<String, String> producer1;
    private final Properties props = new Properties();
    Gson gson = new Gson();

    // Needed string to store the update
    // layout
    private String json;

    // Holding the extracted time stamp build the final update layout
    String t_stmp;

    public NewProducer(String car1, String canList) {
        // update hand-over
        this.json = car1;
        (...)
    }

    private void parseJSelem(JsonElement jsElement) {

        // Three distinctions within the parsing tree JsonObject, JsonArray,
        // JsonPrimitive
        if (jsElement.isJsonObject()) {
            JsonObject jsObject = jsElement.getAsJsonObject();
            Iterator<Map.Entry<String, JsonElement>> it = jsObject.entrySet()
                    .iterator();

            while (it.hasNext()) {
            (...###Parser###...)
                                produceMessage2(finalKeyValuePairs, topic2);
                            }
                        }
                    }
            }
        }
    }
```

As a final step a consumer thread will be started below.

```
public class NewConsumer {
    public static void main (String[]args){
    // @param: ConsumerID - could be application name
    // @param: Topic - subscribed topic
    // @param: GroupID - so far I chose the VIN
    String vin1 = "VWGEF9BP2CD054765";
    // Consumer application 1 with name TBD related to vin as groupID
        Consumer consumerThread_1 = new Consumer("Consumer_1",
            "engineSpeed", vin1);
    consumerThread_1.start();
    }
}
```

The consumer class is now sending out the data request related to a consumer group and including a topic representing the required sensor signal.

```
public class Consumer extends Thread {
    // Needed instantiations
    private final ConsumerConnector consumer;
    // Needed strings to store the topic
    // and consumerID used for the Consumer configuration
    private final String topic;
    private String consumerID;
    (...)
    public void run() {
        // Creating a map of (topic, #streams) pair
        // looking like:{DisplayedVehicleSpeed=1}
        Map<String, Integer> topicCountMap = new HashMap<String, Integer>();
        // setting the first topic as topic one
        topicCountMap.put(topic, new Integer(1));
        // Creating a map of (topic, list of KafkaStream) pairs.
        // The number of items in the list is #streams.
        // Each stream supports an iterator over message/metadata pairs.
        Map<String, List<KafkaStream<Message>>> consumerMap = consumer
                .createMessageStreams(topicCountMap);
        // Extracting the stream to a chosen topic from the consumerMap
        KafkaStream<Message> stream = consumerMap.get(topic).get(0);
        // Creating an iterator over messages in the stream.
        // Allows to extract all the messages from the KafkaStream
        ConsumerIterator<Message> it = stream.iterator();

        while (it.hasNext()) {
            // Displaying the consumer ID with the all the stream messages
            // extracted from the ByteBuffer as Strings
            System.out.println(this.consumerID + ": "
                    + ExampleUtils.getMessage(it.next().message()));
        }
    }
}
```

CHAPTER 6 – EXPERIMENTAL SETUP

This describes the test setup for the framework implementation, as well as the already mentioned Golden Rules in order to evaluate the privacy concept in the subsequent chapter.

IMPLEMENTATION TEST CASES

Test cases for the framework have been defined based on standard black-box and white-box tests, looking at inputs and outputs, and whether the internal data processing operates as required. The system has been tested with two types of sensor data, a common gas vehicle and an electric vehicle (EV). The test data provided in this section is based on the EV data. The source data available comprises roughly 140 sensor data groups with sub-sensors ranging from 2-130. For the following test scenario the available sub-sensors range from 0-3, which also considers empty groups. The conceptual overview of the system below illustrates the inputs and outputs as well the data translation inside the system (Fig.19).

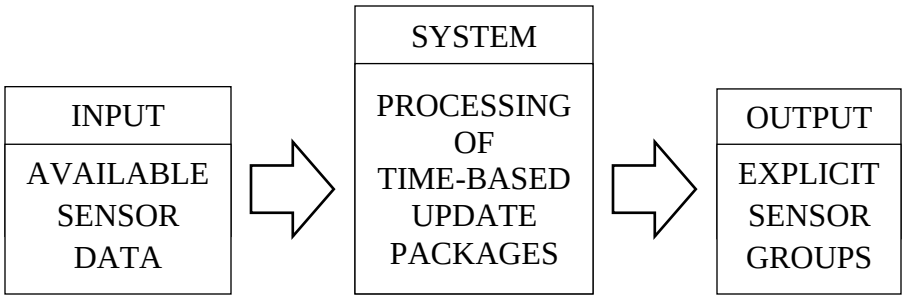

Figure 19: Conceptual system overview

In the following test cases are defined based on standard black-box and white-box criteria. These cases shall be used in the evaluation to test the final implementation of the presented framework.

This first table defines the specific test cases including *name, testing method, and brief description* (Tab.10).

Table 11: Defined Test Cases

Name	***Method***	***Description***
Data set	Creating outputs allowing to compare the original XML file (extracted vehicle data) with newly created JSON file by comparing elements randomly	Checks for the correct format of the testing data defining the vehicles update set. *Successful, if JSON elements match XML elements.*
Update	Creating an output that indicates a successfully received update set including a comparison of data that was sent with the data that has been received	Verifies that the update has been received correctly. *Successful, if sent elements match received elements and prompt indicates update received.*
Producer	Creating an output of the extracted elements that shall be made available and compare them with the original elements from the source file	Verifies that the producer has extracted the vehicle data correctly from the update set. *Successful, if extracted elements match original elements.*

Consumer	Creating an output that shows the subscription which can be validated in two steps. First, the requested data need to match the topic that has been subscribed to. Second, compare the received data with the original file in order to approve the data correctness.	Verifies that the subscriber has received the vehicle correctly from the publisher/producer. *Successful, if subscribed topic matches the prompt indicating the received data. And the received value of the data element matches the value from the original file.*

This second defines the specific inputs and expected outputs corresponding to the test cases (Tab.11).

Table 12: Defined Inputs and Outputs

Name	***Input***	***Expected Output***
Data set	XML file	JSON file
Update	JSON file	JSON file, reception approval
Producer	JSON file	Extracted vehicle data
Subscriber	Consumer ID, group ID, subscribed topic	Vehicle data corresponding to topic, with time stamp of creation

GOLDEN RULES

As part of the Federal Data Protection Act the so called “Golden rules” are mounted into the German, British and most other European countries federal laws. The act focuses on the protection against misuse of personal data in terms of data processing. The definition can e.g. be found as annex to section 9 in the German “Federal Data Protection Act”, which is related to technical and organizational measures and therefore fundamental policy for big companies, like Volkswagen dealing with various amounts of data.

The following stated 10 paragraphs are extracted from the German “Federal Data Protection Act”, German orig.: “Bundesdatenschutzgesetz”.

“Where personal data are processed automatically, measures suited to the type of personal data to be protected shall be taken

1. to prevent unauthorized persons from gaining access to data processing systems with which personal data are processed (access control),
2. to prevent storage media from being read, copied, modified or removed without, authorization (storage media control),
3. to prevent unauthorized input into the memory and the unauthorized examination, modification or erasure of stored personal data (memory control),
4. to prevent data processing systems from being used by unauthorized persons with the aid of data transmission facilities (user control),
5. to ensure that persons entitled to use a data processing system have access only to the data to which they have a right of access (access control),
6. to ensure that it is possible to check and establish to which bodies personal data can be communicated by means of data transmission facilities (communication control)
7. to ensure that it is possible to check and establish which personal data have been input into data processing systems by whom and at what time (input control),

8. to ensure that, in the case of commissioned processing of personal data, the data are processed strictly in accordance with the instructions of the principal (job control) ,
9. to prevent data from being read, copied, modified or erased without authorization during the transmission of personal data or the transport of storage media (transfer control),
10. to arrange the internal organization of authorities or enterprises in such a way that it meets the specific requirements of data protection (organizational control).” (BDSG, 1990)

CHAPTER 7 – EVALUATION

In this chapter it will be evaluated whether the applied privacy mechanisms for the various concepts are satisfying the data protection requirements stated by the Golden Rules (*Chapter 2 and 6*). Further a brief summary of the back-box and white-box tests regarding the framework implementation shall be given. Eventually, it shall be assessed, whether there is a trade-off between the usability of concept functionality and the level of privacy protection.

TESTING THE COMMUNICATION

This section is based on the in *Chapter 6* defined test cases regarding the main concept (Tab.12). In the following the results of the implementation testing shall be presented.

Table 13: Test cases including success criteria

Name	***Success criteria***
Data set	JSON elements match XML elements.
Update	If sent elements match received elements and prompt indicates update received.
Producer	The extracted elements match original elements.
Consumer	The subscribed topic matches the prompt indicating the received data. And the received value of the data element matches the value from the original file.

For the following test scenario the available sub-sensors range from 0-3, which also considers empty groups. In order to differentiate easier between the different formats, a brief syntax overview shall be provided.

In general The XML file can be identified by the <*text*>*content*</*text*> syntax, whereas the JSON file can be identified by the "name": {"name": "value"} syntax.

Case 1: Data set

The following prompts show parts of the JSON file generated from the original XML vehicle data below (Fig.20 & Fig.21).


```
"energyFlow" : {
    "content" : [null, null, null, null, null],
    "mainMachine" : "IDLE",
    "chargingUnit" : "CHARGE",
    "climateSystem" : "IDLE",
    "timeOfGeneration" : "2013-02-15T21:34:50.534+01:00"
},
"engineSpeed" : {
    "content" : [null, null, null],
    "engineSpeed" : 0,
    "timeOfGeneration" : "2013-02-15T21:33:48.570+01:00"
},
```


Figure 20: JSON representation of the vehicle data


```
<energyFlow>
    <chargingUnit>CHARGE</chargingUnit>
    <climateSystem>IDLE</climateSystem>
    <mainMachine>IDLE</mainMachine>
    <timeOfGeneration>2013-02-15T21:34:50.534+01:00</timeOfGeneration>
</energyFlow>
<engineDisplacement/>
<engineOilLevel/>
<enginePower/>
<engineSpeed>
    <engineSpeed>0.0</engineSpeed>
    <timeOfGeneration>2013-02-15T21:33:48.570+01:00</timeOfGeneration>
</engineSpeed>
```


Figure 21: XML source of the vehicle data

Despite the different formats, both show the same data and values, which indicates that the format conversion was done correctly.

Case 2: Update

The following two prompts show parts of the received update (Fig.22 & Fig.23).


```
UPDATE - raw : {
    "ns2:XFCDDataProviderData" : {
        "rearWindowHeating" : {},
        "engineOilLevel" : {},
        "speedLimit" : {},
        "coolantTemperature" : {},
        "headlampFlasher" : {},
        "doorLockControl" : {},
        "seatBeltLock" : {},
        "horn" : {},
        "rearViewCamera" : {},
        "steerAngle" : {},
        "brakingPressure" : {},
        "numberOfCylinders" : {},
        "cruiseControlSpeedTipSwitch" : {},
        "steeringWheelTorqueIntervention" : {},
        "displayedVehicleSpeed" : {},
        "batteryLevel" : {},
        "hazardWarningSystem" : {},
        "exteriorTemperature" : {},
        "pedalForce" : {},
        "windshieldHeating" : {},
        "tractionControlSystem" : {},
        "fogLight" : {},
        "dashboardIlluminationLevel" : {},
        "parkingBrakeApplied" : {},
        "engineSpeed" : {
            "content" : [null, null, null],
            "engineSpeed" : 0,
            "timeOfGeneration" : "2013-02-15T21:33:48.570+01:00"
        },
        "curbDetection" : {},
```


Figure 22: First part of the update


```
"turnSignalLever" : {},
"energyFlow" : {
    "content" : [null, null, null, null, null],
    "mainMachine" : "IDLE",
    "chargingUnit" : "CHARGE",
    "climateSystem" : "IDLE",
    "timeOfGeneration" : "2013-02-15T21:34:50.534+01:00"
},
"semiautomaticTransmission" : {},
```


Figure 23: Second part of the update

The received data are identical to the data that were sent indicating a correct transfer via HTTP-Post.

Case 3: Producer

The following prompts indicate the extracted data from the parser included in the NewProducer class. The top of the first prompt shows the timestamp. The bottom part of the first prompt (Fig.24) and the second prompt (Fig.25) show the received HTTP-Post data processed by the producer with the implemented JSON parser.

```
Feb 26, 2013 2:56:56 PM org.restlet.engine.log.LogFilter afterHandle
INFO: 2013-02-26        14:56:56        192.168.190.135 -       192.168.190.139 8282    POST    /ServerRestletTest/hello

                    Signal Group: rearWindowHeating
                    Signal Group: engineOilLevel
                    Signal Group: speedLimit
                    Signal Group: coolantTemperature
                    Signal Group: headlampFlasher
                    Signal Group: doorLockControl
                    Signal Group: seatBeltLock
                    Signal Group: horn
                    Signal Group: rearViewCamera
                    Signal Group: steerAngle
                    Signal Group: brakingPressure
                    Signal Group: numberOfCylinders
                    Signal Group: cruiseControlSpeedTipSwitch
                    Signal Group: steeringWheelTorqueIntervention
                    Signal Group: displayedVehicleSpeed
                    Signal Group: batteryLevel
                    Signal Group: hazardWarningSystem
                    Signal Group: exteriorTemperature
                    Signal Group: pedalForce
                    Signal Group: windshieldHeating
                    Signal Group: tractionControlSystem
                    Signal Group: fogLight
                    Signal Group: dashboardIlluminationLevel
                    Signal Group: parkingBrakeApplied
                    Signal Group: engineSpeed
                    Element: engineSpeed: 0
```

Figure 24: Timestamp and extracted vehicle data

```
                    Signal Group: energyFlow
                    Element: mainMachine: "IDLE"
                    Element: chargingUnit: "CHARGE"
                    Element: climateSystem: "IDLE"
                    Signal Group: semiautomaticTransmission
                    Signal Group: tirePressure
```

Figure 25: Extracted vehicle data

As the two prompts (Fig.24 & Fig.25) show the same information as the original vehicle data XML (XF), it indicates that the processing was done correctly.

Case 4: Consumer

The following prompts show two different subscriptions of the same implemented consumer (Fig.26 & Fig.27) including the subscription and the corresponding received data.

```
// Consumer application 1 with name TBD related to vin as groupID
    Consumer consumerThread_1 = new Consumer("Consumer_45",
        "engineSpeed", vin1);
consumerThread_1.start();
}

Consumer_45: 2013-02-15T21:33:48.570+01:00: engineSpeed: 0
```

Figure 26: First subscription

```
// Consumer application 1 with name TBD related to vin as groupID
    Consumer consumerThread_1 = new Consumer("Consumer_45",
        "energyFlow", vin1);
consumerThread_1.start();
}

Consumer_45: 2013-02-15T21:34:50.534+01:00: mainMachine: "IDLE"
Consumer_45: 2013-02-15T21:34:50.534+01:00: chargingUnit: "CHARGE"
Consumer_45: 2013-02-15T21:34:50.534+01:00: climateSystem: "IDLE"
```

Figure 27: Second subscription

As both prompts (Fig.26 & Fig.27) show that the consumed data corresponds to the subscribed signal group, it indicates the subscription process was done successfully.

DISCUSSING THE RESPONSIBILTY OF PRIVACY PROTECTION

As it was cited earlier in this thesis, the "Golden Rules" as part of the German "Federal Data Protection Act" are focusing on the protection against misuse of personal data in data processing. It is related to technical and organizational measures and therefore fundamental policy for big OEMs (car manufacturers) like Volkswagen dealing with various amounts of data.

The following tables and graphics will illustrate how the responsibility for privacy protection or often referred to as data protection is divided throughout the lifecycle of a product in an automotive environment. It shall become apparent that not all responsibility relies on the concept but a significant amount of protection mechanisms are dictated by law and therefore already implemented by other parties of the protection process. While indicating what privacy mechanisms are already taken care of by other parties like e.g. the OEM and the developer pending protection needs covered by the concept shall be pointed out as well (Tab.13).

Table 14: Policies and Responsibilities

Policy section	***Responsibility***
1. Access control (unauthorized persons)	OEM, developer
2. Storage media control	Privacy concept
3. Media control	Privacy concept
4. User control	OEM, developer
5. Access control (access rights)	OEM

6. Communication control	Privacy concept
7. Input control	Privacy concept
8. Job control	OEM, developer
9. Transfer control	Privacy concept
10. Organizational control	OEM

There are three major parties, the OEM (car manufacturer), the developer, and the developed concept or later product. The process of data protection can be described as a layered model (Fig.28).

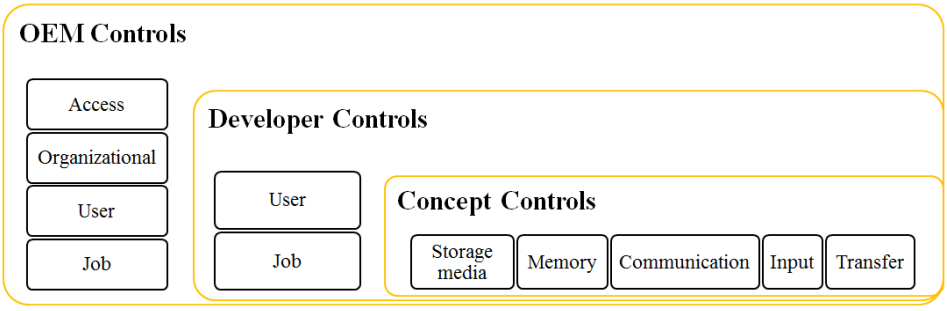

Figure 28: Data protection responsibility breakdown

The outside layer is represented by the OEM covering tasks like access control, communication controls and offering user and job controls. The developer in turn must make use of these provided controls to keep data protected. In the last step, for the concept and later product the developer must incorporate mainly input, transfer and communication protection mechanisms but also make use of secure storage and memory to protect information. The OEM in other words provides the secure environment with right policies, security zones, secure login, and hardware and software encryption for mobile devices (Tab.14, Tab.15). The developer has to make sure that throughout the development phase and later on in the lifecycle the concept implementation allows for data protection. General mechanisms are i.e. choosing

passwords with high security level, using encrypted mobile devices that are accessible to authorized persons only (Tab.14, Tab.15). Further the developer is responsible for the integration of protection mechanisms into the concept and later product.

In the next section it shall be discussed whether the defined privacy requirements agree with the consent of the Golden Rules policies.

PRIVACY PROTECTION MECHANISMS APPLIED TO THE MAIN AND THE RELATED CONCEPTS

Fulfill the privacy protection mechanisms the required policies?

The concept and later product in this context shall be defined as an information exchange system mostly based on software that interacts over mobile Internet connections like e.g. UMTS communicating over powerful antennas covering most of the automotive infrastructure. Therefore i.e. the communication and input controls must be covered by the concept in terms of privacy profiles and auditing mechanisms for transparency purposes (Tab.14). One can define the privacy concept again as a multi-layered concept. The first layer is addressed by the profile preferences that are making sure that only authorized requests are processed and only permitted data leave the vehicle. The decision power clearly remains with the source. The next layer is responsible for the transfer of personal data from the source to the service providing OEM. Private data must be protected through ciphering and key management, making sure that the access is limited again to authorized persons only (Tab.14). The secret key agreement between client and OEM limits the data access additionally. Digital

signatures eventually make sure that only authenticated is considered in order to increase the trust between the parties.

Table 15: Policies and mechanisms regarding the new framework

Policy description (*)	***Mechanisms applied to this policy***
1. "to prevent unauthorized persons from gaining access to data processing systems with which personal data are processed (access control)"	Policies, security zones, public-key infrastructure (PKI) access control, network access control, operating system access control, secure logon procedure, password security, session time out detection, define user responsibilities, offer hardware or software encryption mobile devices (**)
2. "to prevent storage media from being read, copied, modified or removed without authorization (storage media control)"	Data Encryption Standard, RSA Digital Signature
3. "to prevent unauthorized input into the memory and the unauthorized examination, modification or erasure of stored personal data (memory control)"	Data Encryption Standard, RSA Digital Signature
4. "to prevent data processing systems from being used by unauthorized persons with the aid of data transmission facilities (user control)"	Password security, user responsibilities, use hardware or software encrypted mobile devices (**)

5. "to ensure that persons entitled to use a data processing system have access only to the data to which they have a right of access (access control)"	Public-key infrastructure (PKI) access control, secret key management, access right profiles, access on a "need to know" basis (**)
6. "to ensure that it is possible to check and establish to which bodies personal data can be communicated by means of data transmission facilities (communication control)"	Privacy profile permissions, auditing (**)
7. "to ensure that it is possible to check and establish which personal data have been input into data processing systems by whom and at what time (input control)"	Privacy profile permissions, auditing (**)
8. "to ensure that, in the case of commissioned processing of personal data, the data are processed strictly in accordance with the instructions of the principal (job control)"	Authentication, authorization, auditing, job responsibilities (must handle information with care, must apply valid IT security regulations), "doing the right thing" (**)
9. "to prevent data from being read, copied, modified or erased without authorization during the transmission of personal data or the transport of storage media (transfer control)"	Privacy profile permissions, Data Encryption Standard, RSA Digital Signature

10. “to arrange the internal organization of authorities or enterprises in such a way that it meets the specific requirements of data protection (organizational control)”	Responsibilities are defined within the company, protective measures instructed to employees, the information and data security has an advising role and needs to give approval in case of possible exceptions (**)(***)

(*) (BDGS, 1994), (**)(Fröhlich, 2003-2010), (***)(Knerlein, 2002)

PRIVATE HOTSPOT IDENTIFICATION AS A DIFFERENT DATA DISTRIBUTION CONCEPT

Fulfill the privacy protection mechanisms the defined policies?

As in the section before, the “Golden Rules” shall be utilized as a standardized compilation of privacy requirements (Tab.15). As the hotspot identification protocol describes a private interaction between client and OEM, such that no personal data are revealed or stored, the emerging anonymous results therefore only need to be authenticated. In order to prevent tampering with voting results, additional hotspot information as defined by Raghunathan et al. as auxiliary data, shall be protected as well. Accordingly the standard OEM data protection controls remain necessary (Tab.15).

The private hotspot identification protocol describes an alternative data exchange concept based upon a significantly reduced amount of exchanged information. One can describe this concept as privacy protection in different stages. Again, it is mostly

based on software that interacts over mobile Internet connections. Adjusted to the different concept, the communication and input controls are covered by respectively different data protection mechanisms. Besides the necessary auditing instrument for transparency purposes, the concept comprises four main mechanisms.

In order to maintain location privacy, all personal information i.e. identification (ID), password, and actual location are hidden or remain undistinguishable. This is achieved through a commitment scheme that binds to an ID-password pair during the registration phase, while hiding it during the voting phase in order to keep privacy protected (see *Chapter 2*). It is in the responsibility of the individual to not reveal or to even sell this pair to a third party. In case the individual decides to do so anyway, the assumption is that the amount of tampered votes represents an obvious minority that does not allow manipulating the overall voting result. During the voting phase the protocol uses Blind Signatures in order to prevent a linking possibility between the client's signature and account information. This mechanism shall inhibit the identification of a client by his digital signature and thereby enhance anonymity. The third form of anonymity enhancement is achieved through a so called Zero Knowledge Proof of Knowledge (ZKPoK) that is used also during the voting phase. It enables the OEM to validate the clients' votes with zero knowledge about the individual client's account information. In other words the incentive is that the OEM is able to authenticate a client without knowing the actual identity. Additionally, based on the informal proof given in *Chapter 4*, stating that due to the neighboring relation between clients and the rounded location hotspot identification, the OEM cannot determine the actual location of the client.

Eventually, the decision power remains with the source once again. The client can decide if and what to vote for while not revealing any personal information. Although private data are not revealed, several mechanisms work on different stages all relying on each other, it is still good practice to have multilayer security. In order to provide multilayer security the protocol incorporates standard key management between client and OEM limiting the access to protocol data. Digital signatures ultimately make sure that only authenticated data are considered in order to increase the trust between the parties.

Table 16: Policies and mechanisms regarding the hotspot protocol

Policy description (*)	***Mechanisms applied to this policy***
1. "to prevent unauthorized persons from gaining access to data processing systems with which personal data are processed (access control)"	Policies, security zones, public-key infrastructure (PKI) access control, network access control, operating system access control, secure logon procedure, password security, session time out detection, define user responsibilities, offer hardware or software encryption mobile devices (**)
2. "to prevent storage media from being read, copied, modified or removed without authorization (storage media control)"	Commitment Scheme, Blind Signature for registration data. Standard hardware or software encryption is sufficient to protect voting results

3. "to prevent unauthorized input into the memory and the unauthorized examination, modification or erasure of stored personal data (memory control)"	Commitment Scheme, Blind Signature Scheme during registration otherwise no personal information is revealed
4. "to prevent data processing systems from being used by unauthorized persons with the aid of data transmission facilities (user control)"	Password security, user responsibilities, use hardware or software encrypted mobile devices (**)
5. "to ensure that persons entitled to use a data processing system have access only to the data to which they have a right of access (access control)"	Public-key infrastructure (PKI) access control, secret key management, access right profiles, access on a "need to know" basis (**)
6. "to ensure that it is possible to check and establish to which bodies personal data can be communicated by means of data transmission facilities (communication control)"	Commitment Scheme (during registration), auditing
7. "to ensure that it is possible to check and establish which personal data have been input into data processing systems by whom and at what time (input control)"	Only personal is registration data. During the actual voting no personal data are revealed at any time

8. "to ensure that, in the case of commissioned processing of personal data, the data are processed strictly in accordance with the instructions of the principal (job control)"	Authentication, authorization, auditing, job responsibilities (must handle information with care, must apply valid IT security regulations), "doing the right thing" (**)
9. "to prevent data from being read, copied, modified or erased without authorization during the transmission of personal data or the transport of storage media (transfer control)"	Commitment Scheme, Blind Signature Scheme, Zero Knowledge Proof of Knowledge - ZKPoK
10. "to arrange the internal organization of authorities or enterprises in such a way that it meets the specific requirements of data protection (organizational control)"	Responsibilities are defined within the company, protective measures instructed to employees, the information and data security has an advising role and needs to give approval in case of possible exceptions (**)(***)

(*) (BDGS, 1994), (**)(Fröhlich, 2003-2010), (***)(Knerlein, 2002)

DERIVED DATA PRIVACY CLASSES

The data applications investigated in Chapter 2 have been evaluated according to the data that are exchanged. During this privacy-based evaluation, three data categories have been observed: 1) identifiable, 2) time/location based, and 3) broadcast or independent. Table 5 in *Chapter 2* presents the details of observed data categories in various application types. The security classes listed below apply the following four mechanisms, derived from the Golden Rules (*Chapter 6).*

1. Access control mechanisms and rights
2. Secured storage
3. Secured communication
4. Degree of relation to a specific individual

Table 17: Privacy level based on defined data categories

Privacy Classes	***Identifiable data***	***Time/location based data***	***Broadcasted data, independent data***
Personal Data	Name, bank information, social security, insurance information, address, telephone number, vehicle identification number, account information, license plate sensor data, social data, *profile data*		

Personal Related Data		Timed position, driving data, traveled destinations, *Recommendations (routes, charging), Profile data*	
Derivative Data			Preferences (routes, driving style, charging history, news, sport and music), social data
Unrelated Data			Unrelated sensor data, car make, model

The derived data privacy classes define the foundation for the profile management (Tab.16). In general, the classification presented here can be used as guidelines when designing privacy protection mechanisms for the existing and the applications to be.

CHAPTER 8 – RELATED WORK

This chapter will give an overview of the previous attempts to improve privacy protection, as well as introduce currently discussed legal issues.

PRIVACY PROTECTION THROUGH ANONYMITY

Wiedersheim et al. have demonstrated the possibility of reconstructing long traces of a majority of vehicles within the same area. According to their work it is more than questionable if location privacy is achievable in IVC systems against a powerful adversary. Even though actual identities are replaced by pseudonyms and those also change over time, once a target is identified based on its location every vehicle can be tracked. Multi-hypothesis tracking (MHT) is largely recognized as the ideal methodology to solve data association problems in present multiple target tracking (MTT) systems (Blackman, 2004).

The general approach is based on multiple moving targets in a defined area while their position is being sampled at random or periodic intervals. With measurements comes noise and errors. For that reason most MHT systems are combined with filter operations like the Kalman filter. The Kalman filter makes its decision based on the current state, a prediction and a current measurement. The prediction results from known movements of targeted object. The measurement is the outcome of measured beacons (identification messages). The iterative process of prediction is used to

estimate the next position of a vehicle within context of multi hypothesis tracking (Fig.29).

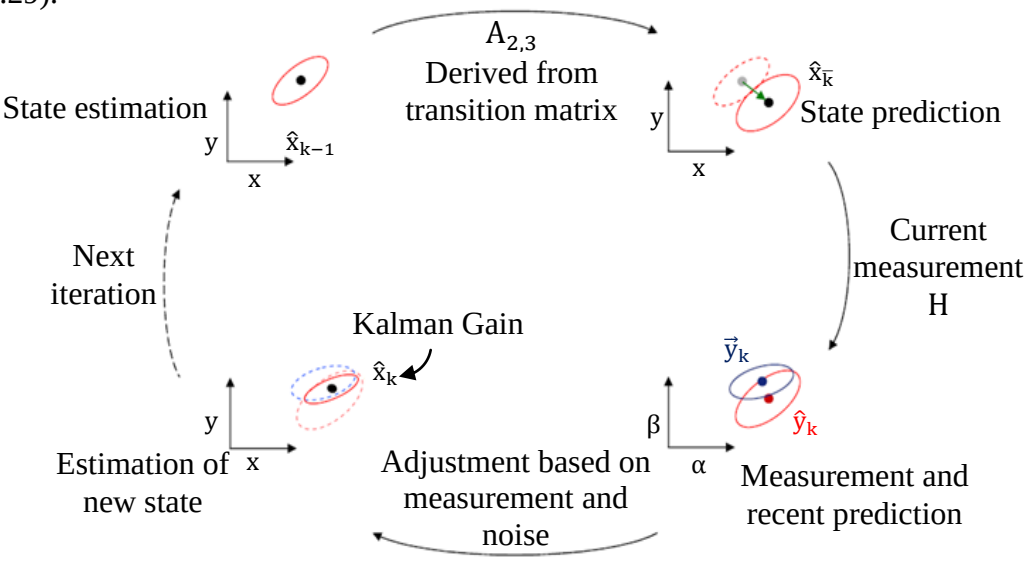

Figure 29: Kalman iterations

In general for each measurement (marked as dots, see Fig.30) one hypothesis shall be created.

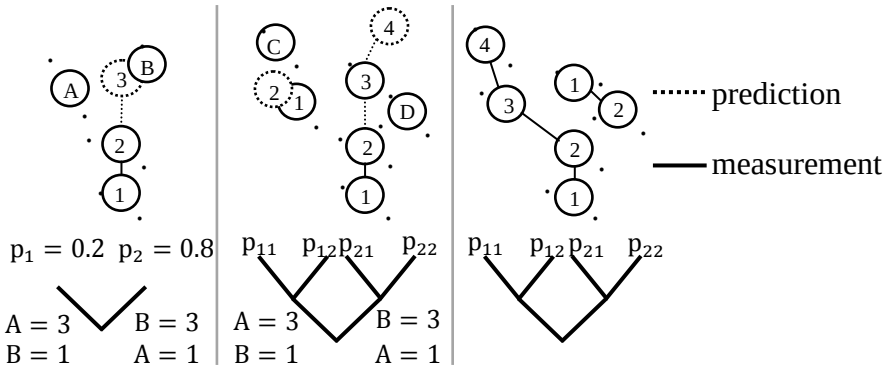

Figure 30: Kalman based muli hypothesis tracking (Wiedersheim et al., 2010)

Predictions are rated with higher probabilities, when they are close to measurements. A hypothesis is defined as one potential track based on a set of measurements. The most likely track is generated from multiple hypotheses.

Eventually the authors show in their evaluation how identities become negligible with traceability. The approach follows the idea of reducing the communication density in terms of adjusting the beaconing (identification messages) intervals (Fig.31).

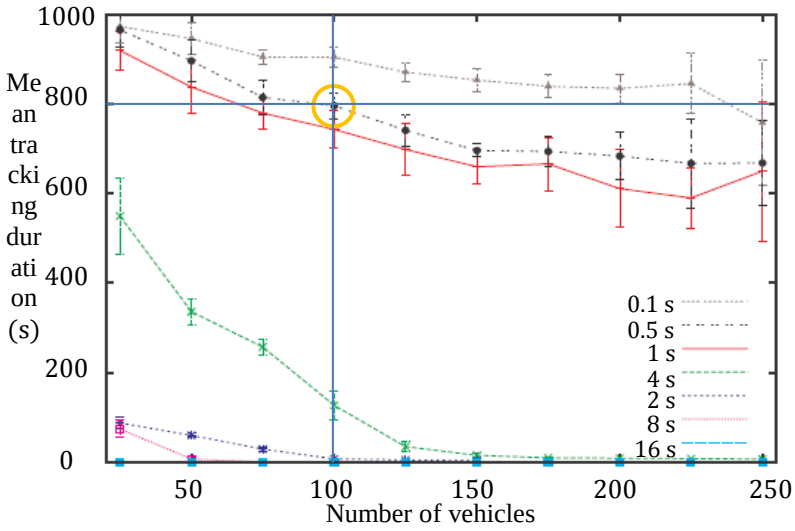

Figure 31: Variation of beaconing intervals (Wiedersheim et al., 2010)

As the graph indicates with a density of 100 vehicles and a beacon interval of 1s or shorter, every vehicle can be tracked for about 800ms out of 1000ms. It seems obvious that decreasing the beaconing rate can be tool to reduce the traceability. Unfortunately $1-10$Hz is commonly discussed in standardization activities to guarantee reliable communication (Schoch et al., 2006). In a second step the authors look into adjusting the pseudonym changing intervals, with the result of necessary pseudonym changing intervals significantly than shorter 30sec. These rates are considered high rates and cause a substantial decrease in performance (Schoch et al., 2006). In other words, both adjustments to enhance privacy are not practical.

Some years ago IBM developed a concept called MyPrivacy Component Architecture (Fig.32). It represents a complete concept presenting a flexible and transparent privacy protection profile that was considered for Internet use only (Bohrer et al., 2001).

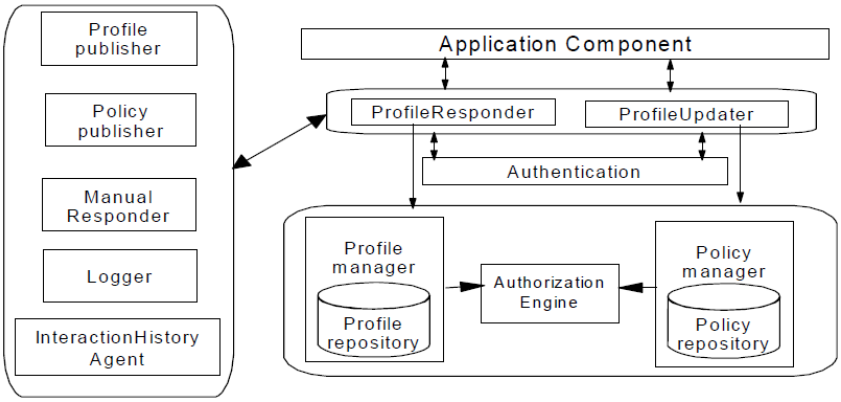

Figure 32: IBM MyPrivacy Component Architecture (Bohrer et al., 2001)

The several components of this architecture shall protect private data. The main functionalities provided by this architecture are profile and policy/rule management accompanied by the authorization engine, as well as the profile responder and profile updater. The primary task of both management components is profile and policy maintenance of the system. The engine shall handle data requests based on the defined policies. The profile responder adds personal privacy statement according to the profile. The profile updater allows for modification within the profile. In case of more complex requests a smaller portion of the requests can be handled manually by the manual components. For more protection the interaction history agent intervenes in case certain actions must be intercepted.

The general idea shall not only be to protect private data but also provide more transparency and scalability when it comes to exposing private data. Instead of agreeing to general business terms specific requests shall be visible in order to understand the purpose and action of the provided services. Since a lot of services out on the market come with reasonable data requests and very useful everyday amenities it should be up to the client to differentiate. The decision shall be service-wise by setting different privacy rules based on the various services (e.g. dealer profiles, workshop profile, tolling system profile).

LATEST PRIVACY DISCUSSIONS

This section will give an abstract on the privacy concerns that are currently discussed in public and in general between the US and EU governments.

The right to be forgotten

The current EU privacy debate seeks to enhance the privacy of its citizen. The declared goal is to strengthen the data protection laws for the web, which is getting slammed by large Internet firms and lobbyists. What happened was that the European Commissioner for Justice Viviane Reding presented a draft (Reding draft) regarding a revised data protection regulation early 2012 as an adjustment to the Internet age. The right for privacy often referred to as "the right to be forgotten" shall arise again to protect the consumers' personal data. The main parties involved in this debate are companies, civil rights proponents and data protection officials in the EU member states. The biggest challenge in data protection is the outcome of the connected world enabling various data combinations. The mainly agreed upon approach is that personal data can be obtained by firms and applications, if they have obtained the consent of the user. The Reding draft specifically states, "the legitimate interests pursued by" the entity that processes the data may make consent unnecessary, as long as such interests are not "overridden by the interests or fundamental right and freedoms of the data subject". In other words, in case the individual representing the "data subject" must be given the opportunity to deny the processing of the data.

Current privacy cases regarding traveling safety

In the last days of 2012 the U.S. National Highway Traffic Safety Administration proposed regulations asking car manufacturers to build event data recorders also known as "black boxes" into all new cars and light trucks (Lowy, J., 2012). The proposal seems far behind since automakers have already been quietly integrating the devices, which are automatically recording the actions of drivers and the responses of their vehicles in a continuous information loop, into most new cars for years. One of the purposes is in case a vehicle is involved in a crash or when its airbags deploy, inputs from the vehicle's sensors within the 5 to 10 seconds before impact are automatically preserved. Within current discussions privacy advocates claim that government regulators and automakers are distributing an intrusive technology without having any policy in place that prevents abuse of the collected data. Some manufacturers already are collecting such data. According to the associate director of the Electronic Privacy Information Center there are no rules or limits, no consequences and there is no transparency. One major concern is that the growing computerization of vehicles and transmission of data to and from vehicles might lead to illegal uses of recorder data.

Current privacy cases regarding social media and connected applications

Social media platforms as they are already implemented into the latest in-vehicle media systems have grown into a major business for all kinds of data trading. Platforms like Facebook are known to make their money by selling their website space and other available data to different companies. Therefore it is not surprising that Facebook is constantly fighting against claims that could compromise this business. One of the current claims is comes from one of the German privacy regulators referring to the fact that Facebook denies the use of pseudonyms. Facebook is officially claiming that the main purpose of the requirement is to keep the website secure (Spiegel-WireReports, 2013).

In the past year it has been detected that Apple and other companies have been transferring their users' calendars from their phones to company servers. The outcome of this controversy had the upside that Apple was forced to change its practices and required apps in its store to ask for permission first, before tapping into address books. California Attorney General Kamala Harris achieved an agreement with Amazon, Apple, Google, Microsoft and others aligning the mobile app industry with existing California law. One of the major requirements asks online services that collect personal data to visibly post a privacy rule. On the downside again, by the end of last year the Senate approved an act that reauthorizes the country's warrantless wiretapping program through 2017. It allows the government to electronically spy on citizens' communications with nearly no control until one of the involved parties is believed to be outside the United States. Proposed amendments are demanding for more transparency (San Francisco Chronicle, 2012).

CHAPTER 9 – CONCLUSION

As Bohrer et al. put it appropriately, the question is not anymore “to give or not to give” personal data, since in the modern world there is no way to live outside the digital or connected world. Everybody has private data stored somewhere at some company or institution. Accordingly, the new question must be “how to deal with the given data and how to control the personal data in the future”, as it now has become a problem of business and society.

Most Internet and smart phone users are already experiencing privacy no longer just as a footnote. It is written in big letters, in order to be transparent and thereby gain the trust of the user, who is perhaps using a smart phone app or an online service. The implementation described in this thesis has shown that the connected world we know from the Internet and smart phones has already reached the automobile. People enjoy the benefits of the Internet, like shopping without necessarily leaving the house. Hence, it is perhaps obvious that people would enjoy similar benefits in their vehicle. A vehicle offering general location information, like points of interests (POIs) with just one scroll move, have already been established in the connected automotive world. Extending this approach by offering personalized route or other recommendations or online diagnosis for your vehicle without the need of seeing a mechanic upfront will bring the same benefits into the automotive world.

The other perspective on this topic is the view of the OEM. In times, where data are flowing constantly, customers become anxious about what is going to happen with their personal data. The origin of this concern comes in most cases of human history

from a lack of the understanding due to the data complexity. Successful companies these days have discovered exactly this, that in order to gain the customer's trust, comprehensible privacy terms are extremely important. The IBM-MyPrivacy Component Architecture was implemented for the traditional Internet usage. Instead of just accepting the terms of conditions of a service, this approach makes sure that the requested information agrees with the terms defined by the user itself. In other words, the service cannot request more data than the user allows, as the negotiation example in *Chapter 4* has demonstrated. This thesis has shown how this profile management concept can also be applied to the connected automotive world and what indication it has for privacy protection. Possible future work could discuss whether the user identification should be managed individually by the service providers or by a central instance, similar to latest Google+ approach.

The discussion on privacy responsibilities has revealed how factors like secured communication, secured storage, access control and related rights are divided among the developing entities involved. The data analysis over the most common automotive and related connected applications has been summarized in new data categories representing the new data complexity. Building on these categories, a new generation of automotive data security classes has been defined, which shall serve as a basis for the adopted privacy profile management. One of features of the privacy management system is flexibility, which has been proven to be important, by demonstrating how data complexity and thereby privacy requirements vary with the service.

Overall, it became apparent that the answer to the question of how to control the data is *transparency*. As people enjoy the benefits of communicating vehicles

(UIEvolution, 2012), it is important to distinguish between the various data provided by the user and define a profile management that acts accordingly.

In contrast, Raghunathan et al. have introduced an approach attacking the challenge of keeping location privacy protected, while practicing location-data mining operations. The proposed system identifies location hotspots (like events or other sights), without learning who is present in particular. As a second feature, the paper introduces a recommendation service building on the hotspot detection.

This approach is clearly a step forward in terms of only sharing what is important, while hiding what has no immediate impact on the subject (e.g. personal data). On the other hand, the reason why this is feasible is that the overall message resulting from the protocol is a "quantity statement". Consequently, this makes the inclusion of e.g. personal data, negligible. In the automotive context, quantity statements can be used in several contexts like sight-seeing, limited diagnosis aspects, aftermarket support, or marketing. Overall, this protocol is applicable to every service, where the only key aspect is the location itself and no personalized service is required.

Eventually, current reports have shown how several manufacturers have already been collecting data without the knowledge and/or consent of the customer. Major privacy concerns result from the growing computerization of vehicles and transmission of data to and from vehicles. The fact that no rules, limits nor consequences have yet been determined demands immediate action improving the overall transparency, as presented in this thesis.

BIBLIOGRAPHY

AutoFieldGuide, "A Hitchhiker's Guide To The Telematics Ecosystem," Available at: http://www.autofieldguide.com/articles/a-hitchhiker%27s-guide-to-the-telematics-ecosystem, Mar. 6th, 2013.

AquaLab, Research in Distributed Computing "STRAW (STreetRAndom Waypoint)," Available at: http://www.aqualab.cs.northwestern.edu/resources/9-projects/144-straw-street-random-waypoint-vehicular-mobility-model-for-network-simulations-e-g-car-networks, Feb. 23rd, 2013.

ASA, Automotive Service Association, "Telematics – Past, Present and Future," pp. 5, USA, May, 2008.

Bamberger, K. A., Mulligan, D. K., "Privacy on the Books and on the Ground," Stanford Law Review", Vol. 63, p. 284-95 , Jan. 2011

UC Berkeley Public Law Research Paper No. 1568385

Barak, B. "Lecture 17 - Zero Knowledge Proofs,", Computer Science Department – Princeton University, p. 1, Apr. 5th, 2010.

Blackman, S. S., "Multiple hypothesis tracking for multiple target tracking," Aerospace and Electronic Systems Magazine, IEEE, p. 5-18, Raytheon, CA, USA, Oct., 2004.

BDGS, Bundesdatenschutzgesetz (Federal Data protection Act), published on Dec. 20, 1990 (BGBl.I 1990 S.2954), as amended by the law (BGBl. I S. 2325), Sep 14th, 1994.

Bohrer, K. and Liu, X. and Kesdogan, D. and Schonberg, E. and Singh, M. and Spraragen, S. L., "Personal Information Management and Distribution," Fourth International Conference on Electronic Commerce Research (ICECR-4), pp. 5, Dallas, TX, USA, Nov., 2001.

Code.google, "dps-x509 – Appendix PublicKeyCryptography, PKI, DigitalSignature, Authentication, Verification" Available at: http://code.google.com/p/dps-x509/wiki/Appendix, Mar. 4th, 2013.

Chaum D. and Pedersen T P., "Wallet databases with observers," In CRYPTO, p. 89–105, 1992.

Chaum, D., "Blind signatures for untraceable payments," Advances in Cryptology - Crypto '82, Springer-Verlag p.199-203, 1983.

Chaum, D., "Security without identification: transaction systems to make big brother obsolete," Communications of the ACM. Oct. 28th, 1985.

Clifton, C. and Jiang, W. and Muruguesan, M. and Nergiz, M. E., "Is Privacy Still an Issue for Data Mining?," National Science Foundation Symposium on Next Generation, Oct. 10-12th, Baltimore, MD, 2007.

Cognizant, "The New Auto Insurance Ecosystem: Telematics, Mobility and the Connected Car," pp. 3, USA, Aug, 2012.

Cordis, "History of the Deployment of transport Telematics," Available at: http://cordis.europa.eu/telematics/tap_transport/intro/benefits/history.htm, Mar. 6th, 2013.

Delfs, H. and Knebl H., "Introduction to Cryptography," Springer, ISBN 978-3-540-49243-6, pp. 38, May 1st, 2007.

Data-Processing, "Public Key Encryption," Available at: http://www.data-processing.hk/glossaries/public-key-encryption/, Feb. 27th, 2013.

Determan, L., "Determann's Field Guide to International Data Privacy Law Compliance," EE, ISBN 978-0857932334, p. 25-47, Oct. 31st, 2012.

Dwork, C. and McSherry, F., "Privacy Preserving Data Mining," Available at: http://www.stanford.edu/group/mmds/slides/mcsherry-mmds.pdf, May 4th. 2010.

Electronics-GPS, "How GPS Receivers Work,"

Available at: http://electronics.howstuffworks.com/gadgets/travel/gps2.htm, Mar. 6th, 2013.

Electronics-TCU, "How the Hughes Telematics Device Works," Available at: http://electronics.howstuffworks.com/gadgets/automotive/hughes-telematics-device3.htm, Mar. 6th, 2013.

EPCD, European Parliament and Council Directive 95/46/EC - on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, O.J. (C 93) ("Directive"), Article 25, 1995.

Ethicspoint, "Data Privacy," Available at: http://www.ethicspoint.com/article/data-privacy, Feb 28th, 2013.

EuroLex, "The protection of individuals with regard to the processing of personal data and on the free movement of such data Directive," 95/46/EC of the European Parliament and of the Council of 24 October 1995, Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML , March 3rd, 2013.

Fiaschetti, A., Suraci, V., Priscoli, F.D., "The SHIELD framework: How to control Security, Privacy and Dependability in complex systems," Complexity in Engineering (COMPENG), 2012.

Ferreira, J.C. and Monteiro, V. and Alfonso, J.L., "Data Mining Approach for Range Prediction Electric Vehicle," Conference on Future Automotive Technology – Focus Electromobility, pp.1-15, Mar. 26-27, Munich, Germany, 2012.

FIPS, Federal information processing standards publication, "Data Encryption Standard (DES)", U.S. DEPARTMENT OF COMMERCE/National Institute of Standards and Technology, pp. 8, Realffirmed, Oct. 25th, 1999.

FIPS, Federal information processing standards publication, "Triple Data Encryption Standard (TDEA)", U.S. DEPARTMENT OF COMMERCE/National Institute of Standards and Technology, pp. 10, Gaithersburg, MD, Jan., 2012.

Fricke, H.-C., Krueger, J., "Exlap communication protocol," Volkswagen Group Research, p. 9, Wolfsburg, Germany, Feb 2nd, 2009.

Fröhlich, D., "Information Technology (IT) Security Policy," Audi – Department for Data Protection and Data Security, p.6, Jul. 1st, Ingolstadt, Germany, 2007.

Fröhlich, D., "Information Technology (IT) Security Guidelines for Employees," Audi – Department for Data Protection and Data Security, p.7, Dec. 21st, Ingolstadt, Germany, 2010.

Fröhlich, D., "Information Technology (IT) Security Guidelines for Developers," Audi – Department for Data Protection and Data Security, pp.4, Oct. 20th, Ingolstadt, Germany, 2006.

Fischermann, T. and Hamann, G., "Wer hebt das Datengold?," Available at: http://www.zeit.de/2013/02/Big-Data, Jan 1st. 2013.

Facebook RepPortal, “Building Brands For The Connected World,” Available at: http://fbrep.com/wp/building_brands.pdf, p.10, Feb., 2012.

Giambruno, A. and Milies, C. P. and Sehgal, S. K., “Groups, Rings, and Group Rings,” International Conference, p. 107, July 28-August 2, 2008, Ubatuba, Brazil.

Gregg, J. A., “On Factoring Integers and Evaluating Discrete Logarithms,” Harvard College, p. 54, May 10th, Cambridge, Massachusetts, 2003.

Goldreich, O., “Zero-Knowledge twenty years after its invention,” MIT Computer Science and Artificial Intelligence Laboratory, p. 2, Rehovot, Isreal, Jul 31st, 2002.

Goldwasser, S., Bellare, M., “Lecture Notes on Cryptography,” Department of Computer Science and Engineering, pp. 235, July, 2008.

Harn L., “Public-key cryptosystem design based on factoring and discrete logarithms,” Computers and Digital Techniques, IEE Proceedings, p. 193–195, May, 1994.

HM Government, Her Majesty’s Governement (“Information sharing”), can be obtained from: www.ecm.gov.uk/informationsharing, Oct., 2008.

Lowy, J., “Car Black Boxes Raise Privacy Concerns,” Huffington Post Dec. 2012, Available at: http://www.huffingtonpost.com/2012/12/07/car-black-box_n_2255110.html, Mar. 3rd, 2013.

Holfelder, W., "Vehicle-to-Vehicle and Vehicle-to-Infrastructure Communication Recent Developments, Opportunities and Challenges,” DaimlerChrysler Research and Technology North America, Inc., Automotive Software Workshop, p. 1, Palo Alto, CA , Jan. 10-12, 2004.

Hustinx, P.J., Opinion 2/99 on the Adequacy of the "International Safe Harbor Principles" issued by the U.S. Department of Commerce on 19th April 1999 - WP 19 (3.051999), May 3rd, 1999.

ISO 15765-2, International Organization for Standardization. ISO 15765-2 Road vehicles - Diagnostics on Controller Area Networks (CAN) - Part 2: Network layer services, 2004.

IMFSurvery Magazine, "The Global Village: Connected World Drives Economic Shift," Available at: http://www.imf.org/external/pubs/ft/survey/so/2012/new083012a.htm, Aug. 30th. 2013.

ITBusinessEdge, "Consumer Privacy Insights and Trends – Q2 2012," Available at: http://www.itbusinessedge.com/slideshows/show.aspx?c=96341, Feb. 2nd, 2013.

Intel, "In-Vehicle Infotainment (IVI)," IntelEmbeddedDesignCenter , Available at: http://www.intel.com/content/www/us/en/intelligent-systems/in-vehicle-infotainment/in-vehicle-infotainment-in-car-entertainment-with-intel-inside.html, Feb. 11th, 2013.

Jansons, J., Bogdanovs, N., Ipatovs, A., "Vehicle-to-Infrastructure Communication based on IEEE 802.11g," International Journal of Digital Information and Wireless Communications, p. 46-50, Mar. 8-12th, 2012.

Json, JavaScript Object Notation, "Introducing JSON", Available at: http://www.json.org/, Feb. 25th, 2013.

Johnston W., and McAllister A., "A Transition to Advanced Mathematics," p. 168, June, 2009.

Kafka, Apache Kafka, "A high-throughput distributed messaging system," Available at: http://kafka.apache.org/design.html, Feb. 24th, 2013.

Knerlein, H. "Minimum Standard "Information Protection"," Audi – Department for Data Protection and Data Security, p.1, Jan. 16th, Ingolstadt, Germany, 1999.

Koshy, T., "Elementary number theory with applications," 2nd edition, Academic Press, ISBN 978-0-12-372487-8, p. 346, May 8th, 2007.

Kuner, C., "Beyond Safe Harbor: European Data Protection Law and Electronic Commerce," 35 Int'l Law, p.79, 84, Nov. 2001.

Kowoma, "Position Determination with GPS," Available at: http://www.kowoma.de/en/gps/positioning.htm, Mar. 6th, 2013.

Lischka, K., Stoecker, C., "All you need to know about the EU Privacy Debate," Spiegel Online, Available at: http://www.spiegel.de/international/europe/the-european-union-closes-in-on-data-privacy-legislation-a-877973.html, Mar. 3rd 2013.

Menezes, A. J. and Van Oorschot, P. C. and Vanstone. S. A., "Handbook of Applied Cryptography Fifth Printing," p. 251, Aug, 2001.

MSDN, Microsoft Developer Network, "Publish/Subscribe," Available at: http://msdn.microsoft.com/en-us/library/ff649664.aspx, Feb. 24th, 2013.

MxRelease, "Symmetric Encryption," Available at: http://www.mxrelease.com/security_mxrelease.htm, Feb. 27th, 2013.

National VII Coalition, "What Is VII?,"

Available at http://www.vehicle-infrastructure.org/WhatsVII.htm, Mar. 16th, 2013.

NRC, National Research Council (U.S.), "The Global Positioning System:

A Shared National Asset," National Academy of Public Administration, p. 16, ISBN 0-309-05283-1, may 31st, 1995.

NYMITY, "The Human Factor in Data Protection," Ponemon Institute, Available at: http://www.nymity.com/Free_Privacy_Resources/Previews/ReferencePreview.aspx?guid=fe300426-6281-4728-bff7-3ad7f5ddb2ce, Mar 22nd, 2012.

Ornstein, A., "The Right to Be Forgotten – US Lobbyists Face Off with EU on Data Privacy Proposal,"

Available at: http://www.spiegel.de/international/business/us-government-and-internet-giants-battle-eu-over-data-privacy-proposal-a-861773-druck.html, Mar. 3rd, 2013.

Oxford, Oxford Dictionary, "Privacy,"

Available at: http://oxforddictionaries.com/definition/english/privacy, Feb. 28th, 2013.

Pedersen, T. P., "Non-interactive and information-theoretic secure verifiable secret sharing," In CRYPTO, p. 129–140, 1991.

Raghunathan, A., Chien, P., Boneh, D. "Privately Identifying Location Hotspots," Stanford University, CA, Apr, 2012.

Regan, P., "Legislating Privacy: Technology, Social Values and Public Policy," University of North Carolina Press, p.199, Chapel Hill, 1995.

Restlet, REST - Representational State Transfer, "Framework for the Java platform," Available at: http://www.restlet.org/documentation/2.0/tutorial, Feb. 25th, 2013.

RFC 4158, "Internet X.509 Public Key Infrastructure: Certification Path Building," Network Working Group 2005, Available at: http://tools.ietf.org/html/rfc4158, Mar. 6th, 2013.

RFC 5280 "Internet X.509 Public Key Infrastructure Certificateand Certificate Revocation List (CRL) Profile," Network Working Group 2008, Available at: http://tools.ietf.org/html/rfc5280, Mar. 6th, 2013.

Ristanovic, N., Papadimitratos, P., Theodorakopoulos1, G., Hubaux, J.-P., and Le Boudec, J.-Y., "Adaptive Message Authentication for Multi-Hop Networks," Wireless On-Demand Network Systems and Services – WONS, p. 96 – 103, 26-28 Jan. 2011.

Rivest, R.L. and Shamir, A. and Adleman, L., "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *CACM*, p. 96-99, Jan 26th, 1983.

San Francisco Chronicle, "Looking back on privacy developments" Issue – Dec. 30th, 2012, Available at: http://www.pressdisplay.com/pressdisplay/viewer.aspx, Mar 3rd, 2013.

Scarfone, K., Jansen, W., Tracy, M., "Guide to general Server Security," National Institute of Standard Technology – NIST publication 800123, section 2-4, Gaithersburg, MD, June 2008.

Scheer, D., "For Your Eyes Only – Europe's New High-Tech Role: Playing Privacy Cop to the World — U.S. Companies Run Afoul of EU Laws on Sharing and

Collection of Data — GM's Phone-Book Odyssey," Wall Street Journal, Oct. 10th, 2003.

Schoch, E., Kargl, F., Leinmuller, T., Schlott, S., and Papadimitratos, P., "Impact of Pseudonym Changes on Geographic Routing in VANETs," Security and Privacy in Ad-hoc and Sensor Networks – European Workshop – ESAS, p. 14, Sep. 20-21st, 2006.

Schweppe, D., Roudier, H. Y., "Security and privacy for in-vehicle networks," Vehicular Communications, Sensing, and Computing (VCSC), 2012 IEEE 1st International Workshop, 2012.

Solove, D., Schwartz, P. M., "Information Privacy Law," 4thed, Aspen Casebook Series, ISBN 978-0735510401, p. 915-17, Dec 9th, 2011.

SpiegelOnline, "Privacy vs. Security – EU Eyes Massive Collection of Air Passenger Data," Available at: http://www.spiegel.de/international/europe/european-parliament-to-debate-own-database-for-flight-passengers-a-871953-druck.html, Mar. 3rd, 2013.

Spiegel-WireReports, "Real-Name Policy Under Fire – Privacy Champions Hand Facebook an Ultimatum Data,"

Available at: http://www.spiegel.de/international/germany/german-state-gives-facebook-an-ultimatum-over-real-name-policy-a-873562-druck.html, Mar. 3rd, 2013.

Spiegel-Wires, "Surfing for Details – German Agency to Mine Facebook to Assess Creditworthiness,"

Available at: http://www.spiegel.de/international/germany/german-credit-agency-plans-to-analyze-individual-facebook-pages-a-837539-druck.html, Mar. 3rd, 2013.

Stratford, J.S., Stratford, J., "Data Protection and Privacy in the United States and Europe," p.17, Fall, 1998.

TelematicsUpdate, "Insurance Telematics," Available at: http://analysis.telematicsupdate.com/insurance-telematics/insurance-telematics-and-data-standards, Jan 23rd, 2013.

UIEvolution, "Connected World," UIEvolution News - (CES) 2013, Available at: http://www.uievolution.com/news/category/connectedcar/, Dec 4th, 2012.

Weisstein, E. W., "Totient Function." From MathWorld – A Wolfram Web Resource. Available at: http://mathworld.wolfram.com/TotientFunction.html, Feb 2nd, 2013.

Weisstein, E. W., "Euler's Totient Theorem," From MathWorld – A Wolfram Web Resource. Available at: http://mathworld.wolfram.com/EulersTotientTheorem.html, Feb 2nd, 2013.

Wiedersheim, B., Ma, Z., Kargl, F., and Papadimitratos, P., "Privacy in Inter-Vehicular Networks: Why simple pseudonym change is not enough," Wireless On-demand Network Systems and Services (WONS), p.176 - 183, Feb. 3-5th, 2010.

O'Connor, T.P., "Trends in chemical concentrations in mussels and oysters collected along the U.S. coast from 1986 to 1993," *Marine Environmental Research*, 41, 1996, pp.183-200